University of Alberta

Combining Gel Electrophoresis and Liquid Chromatography with Mass Spectrometry for Proteome Analysis

by

©

Thuy Tien Tran Quach

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of *Master of Science*

Department of *Chemistry*

Edmonton, Alberta
Spring 2004

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-96541-4
Our file Notre référence
ISBN: 0-612-96541-4

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

With sufficient separation of complex protein mixtures, MALDI MS is an important analytical tool for detection of proteins. In this work, two important areas of proteomics research were studied: the detection of hydrophobic membrane proteins and cancer proteomics. Our work is the first demonstration of using sequential CNBr/trypsin digestion for the analysis of low amounts of hydrophobic membrane proteins in-gels. Peptide mass mapping using MALDI MS and sequence information by MALDI MS/MS enabled the confirmation of protein identification. The inherent advantages and limitations of the mass spectrometric technique are discussed and the usefulness of complementary techniques, such as ESI-MS with on-line MS/MS capabilities, is demonstrated. For the cancer proteomics work, separation techniques of polyacrylamide gel electrophoresis and liquid chromatography were combined with MALDI MS. The direct comparisons of protein expression and detection of the differentially expressed proteins in squamous carcinoma cells are demonstrated to be well suited for biomarker discovery.

To my Grandmother (Ba Nguoi),
Mom, Dad, Nguyen, Hieu, Quoc
and Luke

To God be all the glory

Why do you say, O Jacob,
and complain, O Israel,
"My way is hidden from the LORD;
my cause is disregarded by my God"?
Do you not know?
Have you not heard?
The LORD is the everlasting God,
The Creator of the ends of the earth.
He will not grow tired or weary,
and his understanding no one can fathom.
He gives strength to the weary
and increases the power of the weak.
Even youths grow tired and weary,
and young men stumble and fall;
but those who hope in the LORD
will renew their strength.
They will soar on wings like eagles;
they will run and not grow weary,
they will walk and not be faint.

Isaiah 40:27-31

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Liang Li for granting me the opportunity to work with him and for his guidance and constructive discussions throughout my graduate work. His care in delivery of critique and enthusiastic encouragement to press on towards the goal was truly appreciated.

I would like to thank the committee members, Professor Monica Palcic and Professor Manijeh Pasdar for their active participation during my oral examination and for their evaluation and comments regarding this thesis.

To my colleagues at the laboratory, thank you for your warm welcome when I first joined the group, you have been a delight to work with. Special thanks to Ms. Jing Zheng, Dr. Nan Li, Dr. Nancy Zhang, and Dr. Bernd Keller for their practical advice and professional teaching instrument and protocol use. To Dr. Dawn Richards, Dr. Zhengping Wang, Dr. Boyan Zhang, Dr. Rui Chen, Dr. Alan Doucette, Dr. Huaizhi Liu, Mr. David Craft, Ms. Hongying Zhong and Mr. Christopher McDonald for your helpful discussions. A special thank you for Ms. Xinlei Yu, Ms. Lidan Tao, and Mr. Cheng-Jie Ji who joined the Li group the same time as I did. It was a pleasant journey through graduate school together.

I would like to thank the staff at the Department of Chemistry of the University of Alberta for making student life easier. I would like to especially thank Ms. Ilona Baker and Ms. Jeannette Loiselle for all of your administrative help, the paperwork and answering questions.

I would like to thank the Department of Chemistry and the University of Alberta for the opportunity to learn and grow as a researcher and for providing a safe and well-equipped facility for me to conduct my experiments and the Alberta Cancer Board and the National Science and Engineering Research Council of Canada (NSERC) for providing funding for my research project.

I would like to especially thank Ms. MacDonald, my chemistry teacher at Queen Elizabeth High School, Halifax, Nova Scotia for opening my eyes to the world of chemistry and for teaching me in a way that made me love chemistry. I would like to also thank my previous supervisors, Dr. A. Chatt, Dr. R. Guy and Dr. L. Ramaley, during my undergraduate years at Dalhousie University, Halifax, Nova Scotia, for their outstanding teaching, encouragement and guidance.

I would like to thank Dr. Barbara Knoblach and Professor Marek Michalak at the Dept. of Biochemistry, University of Alberta for the preparation of the ER fractions. My sincere thanks to Dr. Knoblach for her helpful pointers and discussion governing the 2D gel electrophoresis work on the tongue skin cancer cells.

Special thanks to Mr. Paul Semchuk, the Institute for Biomolecular Design, University of

Alberta for running the LC-MS/MS on the ER sample.

I would also like to especially thank Dr. Laiji Li and Dr. Manijeh Pasdar, of the Dept. of Cell Biology at the University of Alberta for collaborative work on the cancer cells research. Thanks to Dr. Laiji Li for your labour in preparing quality cancer cells and Dr. Pasdar for your helpful directions and discussions in this exciting area of research.

I would like to thank Dr. Monica Palcic's group for allowing me to use their equipments such as the ultracentrifuge and the spectrophotometer. In particular, thank you Dr. Adam Szpacenko for your help in using the ultracentrifuge and Dr. Sandra Marcus for your help with the French Press and guide in proteolysis prevention.

Life in Edmonton would have been more cold and dry without my friends in Edmonton. Thank you for the fun, laughter, and support and for being my family away from home.

To my family: Mom, Dad, Nguyen, Hieu, Quoc, Duy and Phuc. Thanks for your love, support and encouragement.

I would like to especially thank my sweetie, Dr. Luke Chen, for your unfailing love and encouragement. I will always remember the three years apart and treasure how it made us stronger.

I would like to thank our Father who art in heaven, the LORD almighty, the creator of all things and source of joy, love, and purpose.

Table of Contents

Chapter 1 1

Combining Gel Electrophoresis and Liquid Chromatography with Mass Spectrometry for Proteome Analysis 1

1.1 MALDI TOF Mass Spectrometry 1

1.1.1 MALDI Process 3

1.1.2 Time-of-Flight Mass Analyzer 5

1.1.3 Mass Accuracy, Mass Calibration and Mass Resolution 7

1.1.4 Detector 11

1.2 Introduction to Quadrupole-TOF Mass Analyzer 12

1.2.1 Quadrupole Mass Analyzer 13

1.2.2 Basic Operation of MALDI QqTOF MS and MS/MS 16

1.2.3 Fragmentaton in MS/MS 18

1.3 Separation Methods 22

1.3.1 2D Polyacrylamide Gel Electrophoresis 23

1.3.2 Basic Principles of Isoelectric Focusing 24

1.3.3 Basic Principles of SDS-PAGE 24

1.3.4 RP HPLC 27

1.4 Scope of the Thesis Work 28

1.5 Literature Cited 30

Chapter 2 36

Development and Applications of In-Gel CNBr/Tryptic Digestion Combined with Mass Spectrometry for the Analysis of Membrane Proteins 36

2.1 Introduction 36

2.2 Experimental 39

  2.2.1 Materials and Reagents 39

  2.2.2 Bacteriorhodopsin 40

  2.2.3 NAR I E. Coli 40

  2.2.4 ER Membrane Proteins 41

  2.2.5 In-gel Digestion 42

  2.2.6 Sample Preparation for MALDI MS Analysis 43

  2.2.7 MALDI Mass Spectrometry 44

  2.2.8 MALDI Tandem Mass Spectrometry (MS/MS) 44

  2.2.9 Mass Spectra Interpretation and Database Searching 45

2.3 Results and Discussion 45

  2.3.1 Development of CNBr/Trypsin Digestion Method Using Bacteriorhodopsin 45

  2.3.2 Effect of *n*-Octyl Glucoside on MALDI MS Detection 50

  2.3.3 Detection Sensitivity of CNBr/Tryptic Digest of BR In-Gel 52

  2.3.4 Source of Unrecovered Peptides From a MALDI MS Analysis of an In-Gel CNBr/Tryptic Digest 54

  2.3.5 Application to a Membrane Protein with Cysteine Groups: NAR I E. Coli. 55

  2.3.6 Application to a Complex Biological Extract: ER Membrane Proteins

(Mouse) 57
2.3.7 Identification of ER Membrane Proteins 62
2.4 Conclusions 66
2.5 Literature Cited 67

Chapter 3 69

Mass Spectrometric-Based Approach to Protein Profiling of Squamous Cell Carcinoma of the Human Tongue 69
3.1 Introduction 69
3.1.1 Cancer Proteomics 69
3.1.2 Reduced Cell-Cell Adhesiveness in Cancer Cells 72
3.1.3 Mutations in Cell Adhesion Proteins 73
3.1.4 Squamous Carcinoma Cells 74
3.2 Experimental 76
3.2.1 Materials and Reagents 76
3.2.2 Sample Preparation 77
3.2.3 2D Gel Electrophoresis 79
3.2.3.1 Reproducibility Study of 2D-PAGE Method 79
3.2.3.2 Protein Profiling by 2D-PAGE 80
3.2.4 Direct MALDI MS and RP HPLC Offline MALDI MS 81
3.3 Results and Discussion 83

3.3.1 Protein Profiling by 2D PAGE 83

3.3.1.1 Reproducibility of the 2D PAGE Method 83

3.3.1.2 Proteome Map Generated by 2D PAGE 86

3.3.2 Protein Profiling by Direct MALDI MS and RP HPLC Offline MALDI MS 91

3.3.2.1 Direct MALDI MS 92

3.3.2.2 Reproducibility Study of the HPLC Offline MALDI MS Method 97

3.3.2.3 Reproducibility Study of Extraction Method 102

3.3.2.4 Profiling of Protein Expression in Squamous Carcinoma Cells by RP HPLC offline MALDI MS 108

3.4 Conclusions 115

3.5 Literature Cited 117

Chapter 4 120

Conclusions and Future Work 120

List of Tables

Table 2.1 Identification and Detection of Peptides from Band U7 64

Table 3.1 The cell adhesion proteins expressed in SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue 79

Table 3.2 Identification of proteins showing distinct changes in SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue by trypsin digestion, MALDI TOF MS and MS/MS analysis 90

Table 3.3 Protein profiling of SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue by direct MALDI MS 94

Table 3.4 Reproducibility study of RP HPLC offline MALDI MS of the Human squamous carcinoma cells of the tongue, three repeated runs of the same extract mixture of SCC9 cells 98

Table 3.5 Reproducibility study of RP HPLC offline MALDI MS of the Human squamous carcinoma cells of the tongue, three runs of three different extract mixture of SCC9 cells 105

Table 3.6 Protein profiling of SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue by RP HPLC offline MALDI MS 109

List of Figures

Figure 1.1 Schematic of MALDI-TOF mass spectrometer in linear mode 5

Figure 1.2 Schematic of MALDI-TOF mass spectrometer in reflectron mode 10

Figure 1.3 Schematic of a quadrupole mass spectrometer 13

Figure 1.4 Plot of a-q stability diagram 15

Figure 1.5 Low energy CID fragmentation nomenclature 19

Figure 2.1 Bacteriorhodopsin sequence showing the seven transmembrane regions. The solid underlined sequence of BR shows the cleaved peptide of an in-gel trypsin digestion of 1 μg BR. The dashed underlined sequence of BR shows the cleaved peptide of an in-gel CNBr digestion of 5 μg BR. The dotted underlined sequence of BR shows the cleaved peptides of an in-gel CNBr/trypsin digestion of 5 μg BR 47

Figure 2.2 MALDI mass spectra of BR using various digestion protocols: (A) In-gel trypsin digest of 1 μg BR, (B) In-solution CNBr digest of 20 μL of 1 μg/μL BR, (C) In-gel CNBr digest of 5 μg BR and (D) In-gel CNBr/trypsin digest of 5 μg BR. The peaks labeled with the corresponding m/z value are from the peptides of bacteriorhodopsin. The masses of the peaks shown in the inset in (B) were determined by external calibration. The peak labeled with 'T' is the trypsin peptide (LGEDNINVVEGNEQFISASK) at m/z 2163.05. 48

Figure 2.3 MALDI mass spectra of in-gel CNBr/trypsin digestion of BR: (A) 200 ng BR in-gel and (B) 500 ng BR in-gel using 0.005% n-OG in digestion buffer. (C) MALDI QqTOF MS/MS spectrum of in-gel CNBr/trypsin digestion of 500 ng BR peptide (RPEVASTFK) at m/z 1034.5. (D) MALDI mass spectrum of an in-gel trypsin/CNBr digestion of 500 ng BR. The peaks labeled with M, C, K, and T originate from the HCCA matrix, common contaminate peaks of in-gel experiments, keratin and tryptic peptide peaks, respectively 53

Figure 2.4 (A) Sequence of NAR I showing the penta transmembrane regions with CNBr/trypsin cleaved peptides (solid underline). (B) MALDI mass spectrum of in-gel CNBr/trypsin digestion of 1 μg NAR I. (C) MALDI QqTOF MS/MS spectrum of in-gel CNBr/trypsin digestion of 1 μg NAR I peptide (TLFLLFPFSR) at m/z 1240.8. The peaks labeled with the corresponding m/z value is specific to the peptides of NAR I 56

Figure 2.5 MALDI mass spectra of Band U7 from the protein extract of the ER membrane of mice. (A) In-gel trypsin digestion (B) In-gel CNBr/trypsin digestion 59

Figure 2.6 MALDI QqTOF MS/MS spectrum of the ligase peptide: (A) peptide (ALHDQLGLR) at m/z 1022.5 in the CNBr/trypsin digest of Band U7. ESI QTOF MS/MS spectrum of the ATP-binding protein: (B) peptide (KITEDTVEFGS) at m/z 1225.6 in the trypsin digest of Band U7. (C) peptide (ITELMQVLK) at m/z 1090.6 in the CNBr/trypsin digest of Band U7 60

Figure 3.1 Diagram of the cell adhesion system at the adherens junction 73

Figure 3.2 Reproducibility study of 2D gel electrophoresis of Human squamous carcinoma cells of the tongue 85

Figure 3.3 Protein profiling of SCC9 cells of the Human squamous carcinoma of the tongue by 2D gel electrophoresis MALDI MS 87

Figure 3.4 Protein profiling of SCC9-E-cad cells of the Human squamous carcinoma of the tongue by 2D gel electrophoresis MALDI MS 88

Figure 3.5 Protein profiling of SCC9-Pg cells of the Human squamous carcinoma of the tongue by 2D gel electrophoresis MALDI MS 89

Figure 3.6 Mass spectra of direct MALDI MS of Human squamous carcinoma cells. 95

Figure 3.7 HPLC chromatograms of Human squamous carcinoma cells 96

Figure 3.8 (A) Mass spectra of an HPLC offline MALDI MS experiment of the fraction collected at 24 minutes. (B) The component at m/z 6649 is observed in all three different extractions of SCC9 at times 24 and 25 minutes 103

Figure 3.9 Mass spectra of HPLC offline MALDI MS of SCC9 cells. The component at m/z 10,836 is observed in several fractions, at times 21, 35, and 36 minutes 104

Figure 3.10 Venn diagram of protein profiling of SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue by HPLC offline MALDI MS 112

List of Abbreviations

Aj	Adherens junction
APS	Ammonium persulfate
ATCC	American Type Culture Classification
Bis	N,N'-ethylene-bis-acrylamide
CID	Collision-induced dissociation
dc	Direct current
DTT	Dithiothreitol
E-cad	E-cadherin
ER	Endoplasmic Recticulum
ESI	Electrospray ionization
FWHM	Full width at half-maximum
HCCA	α-cyano-4-hydroxycinnamic acid
IEF	Isoelectric focusing
IPG	Immobilized pH gradient
m/z	Mass-to-charge ratio
MALDI	Matrix-Assisted Laser Desorption Ionization
MCP	Microchannel plate
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
n-OG	*n*-octyl glucoside
pI	Isoelectric point
Pg	Plakoglobin

ppm	Parts-per-million
rf	Radio frequency
RP HPLC	Reversed phase high performance liquid chromatography
SCC9	Squamous Cell Carcinoma Line
SCC9-E-cad	Squamous Cell Carcinoma Line transfected with E-cadherin
SCC9-Pg	Squamous Cell Carcinoma Line transfected with Plakoglobin
SDS	Sodium Dodecyl Sulphate
S/N	Signal-to-noise ratio
TOF	Time-of-Flight
2D-PAGE	Two-Dimensional Polyacrylamide Gel Electrophoresis
UV	Ultraviolet

Chapter 1

Introduction:

Combining Gel Electrophoresis and Liquid Chromatography with Mass Spectrometry for Proteome Analysis

Protein characterization and identification by mass spectrometry (MS) has become the method of choice in proteome analysis. Its popularity is largely due to advances in soft ionization techniques, particularly matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), as well as, advances in bioinformatics search tools after the completion of several genome sequencing projects. These gentle ionization methods made it possible to ionize large thermally labile biomolecules and transfer them to the gas phase without inducing extensive thermal decomposition. Mass spectrometry can be readily combined with separation methods and offers high sensitivity, accuracy, and rapid analysis. However, its application towards proteome research of complex cellular extracts and towards the detection of hydrophobic proteins and low abundance proteins still require improved methods. In the subsequent sections, an overview is given of the MALDI process and the technology and principles underlying the time-of-flight (TOF) and quadrupole-TOF mass spectrometers. As well, a brief summary of the separation techniques of 2D gel electrophoresis and liquid chromatography (LC) will be outlined.

1.1 MALDI TOF Mass Spectrometry

As early as the 1960s, lasers were used in mass spectrometry to generate ions from organic molecules but proved problematic for analysis of high mass compounds, with the limited mass up to 1000 Da for biopolymers.[1] Biopolymer molecules, such as proteins and carbohydrates, normally present in the condensed phase are extremely non-volatile because they are polar and massive, and therefore are difficult to convert into intact, isolated ionized molecules in the gas phase. Many volatization and ionization techniques have been developed[2], including those that generate ions by the application of a high electric field to the sample (field desorption[3]); by bombardment of the sample with energetic ions or atoms (^{252}Cf plasma desorption[4] and secondary ion MS[5]); by the formation of ions directly from small, charged liquid droplets (thermospray ionization[6] and electrospray ionization[7]); and by bombardment with short duration, intense pulses of laser light (laser desorption[8]). Of these techniques, ESI and MALDI, are now widely used for the mass spectrometric analysis of biopolymers. Both the groups of Karas and Hillenkamp[9] and Tanaka et al.[10] in 1987 independently introduced the MALDI technique. Because this is a pulsed technique, it was most easily interfaced with TOF mass spectrometers. As a result, the molecular weight determination of proteins with high masses in excess of 100 kDa was possible.[11,12] In 1989, Fenn and coworkers[13] introduced ESI, commonly used with quadrupole mass spectrometers, as a complementary method of measuring high molecular weights from a series of multiply charged ions.

The use of MALDI TOF MS has become a valuable tool for the structural characterization of proteins. Its popularity stems from the low sample amount required for analysis (sensitivity of femtomoles to picomoles of analyte), its tolerance to buffers and salts, and its high mass accuracy (0.01-0.1%).[2] The following sections will discuss the three major components of a mass spectrometer: ionization, separation and detection. Only the methods/instrumentation relevant to the experiments discussed in this thesis will be outlined.

1.1.1 MALDI Process

In MALDI, the analytes are mixed with a saturated solution of matrix in either pure water or a water/organic (methanol or acetonitrile) solvent. For optimal ion production, a molar ratio of >1000:1 of matrix to analyte should be used. The matrix is a small organic molecule and highly absorbing, typically in the ultraviolet (UV) range. Common matrices used in peptide/protein analysis that can tolerate a relatively high amount of buffer and salts normally used in the preparation of biological samples, are α-cyano-4-hydroxycinnamic acid[14] (HCCA)(<10000 Da) and sinapinic acid (for larger peptides and proteins)[15], while 2,5-dihydroxybenzoic acid (DHB)[16] is good for polar and non-polar samples. Several sample-matrix preparation methods have been developed, including dried-droplet[12], vacuum-drying[17], crushed-crystal[18], slow crystallization[19], fast evaporation,[20,21] sandwich[22,23] and two-layer method[24]. HCCA as matrix and the two-layer method is most commonly used in this laboratory. HCCA has superior detection sensitivity and higher tolerance to salts for protein samples. The two-layer method was shown to be ideal for MALDI MS analysis of samples containing up to 1% SDS.[25]

In the two-layer method, the first layer, consisting of a thin layer of matrix, is deposited on the stainless steel sample plate and is used as a 'seed' for crystallization. The second layer, consisting of the mixture of analyte mixed in a saturated solution of matrix, is then applied on top of the first layer. The solvent evaporates and the matrix and the analytes co-crystallize on the target. Water is used to clean ionic contaminants such as salts on the sample spot and air-dried. The sample is then ready to be loaded into the mass spectrometer and analyzed. The two-layer method produces high resolution mass spectra, better reproducibility from shot to shot and better external mass calibration accuracy afforded by the "homogeneous" co-crystals of matrix/analyte resulting from this deposition method.

Two types of pulsed lasers, UV lasers (nitrogen and Nd:YAG) and IR lasers (TEA-CO_2 and Er:YAG), have been commonly employed in MALDI, with pulse durations between 1-200 ns.[2] A pulsed laser beam, a nitrogen laser at 337 nm and 3 ns pulse used in our experiments, provides light that is absorbed by the aromatic matrix molecules which dissipates the absorbed energy thermally to analyte molecules and becomes desorbed into the gas phase. This rapid transfer of energy causes the matrix to rapidly dissociate from the surface, generating a plume of matrix and the co-crystallized analytes into the gas phase. The dissociation of matrix molecules into an assortment of small, neutral fragments, protonated ions and free-radical ions induces the ionization of analyte molecules. The ionization mechanism is not fully understood and several suggestions are still debated.[26] It is unclear whether the analytes acquire their charge during the desorption process (in the solid phase) and/or after entering the plume of molecules by interacting with the matrix molecules in the gas phase.

The matrix is believed to serve three major functions in MALDI: absorption of energy from the laser light, isolation of the biopolymer molecules from each other and proton transfer. Efficient and controllable energy transfer is ensured and decomposition of the analyte molecules is prevented. As well, the dilution of the sample with the matrix prevents potential association of the analyte molecules with each other, forming complexes that may irreversibly interact with the target surface or be too large to be desorbed and analyzed. MALDI produces mainly singly charged ions, which makes mass assignment easy for complex biological mixtures, such as protein digests.

1.1.2 Time-of-Flight Mass Analyzer

The simplest instrumental set-up of a time-of-flight mass analyzer, operated in linear mode, consists of a short source-extraction region (s, on the order of a few cm), a field-free drift region (D, from 0.5 to 4.0 m in length), and a detector (Figure 1.1).

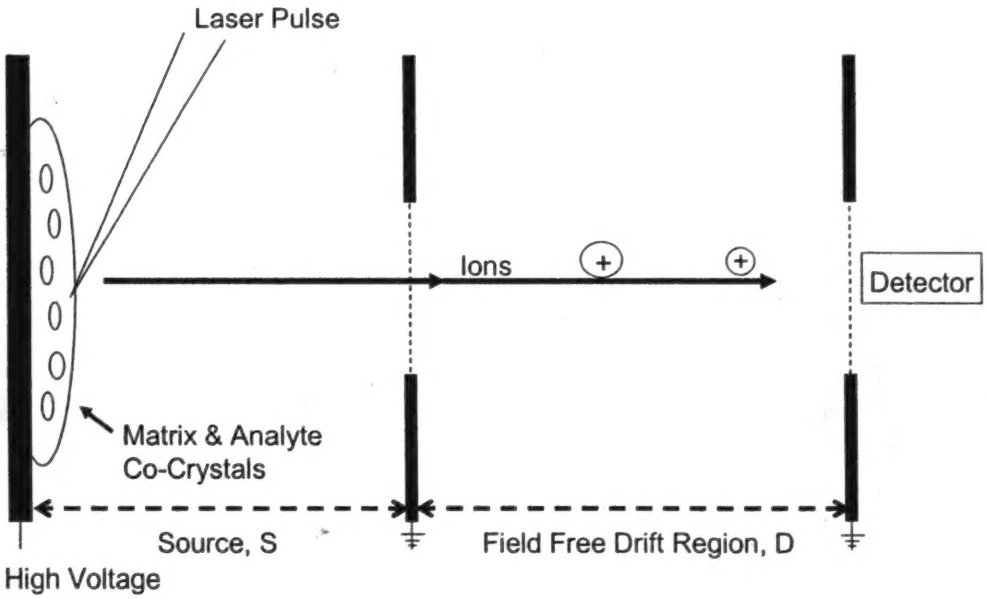

Laser Pulse
Ions
+
+
Detector
Matrix & Analyte Co-Crystals
Source, S
Field Free Drift Region, D
High Voltage

Figure 1.1 Schematic of linear mode TOF.

The principle of the TOF mass analyzer[2,26] is to measure the flight time of ions, starting as they are accelerated out of an ion source into a field-free drift tube, where they are separated according to their mass-to-charge ratio, and ending as they arrive at the detector for signal recording. In MALDI, the ions are formed in the gas phase in the source region. The ions experience an electric field (E=V/s) when a voltage is applied on the source backing plate, which accelerates the ions to constant energy and enter the drift region. The drift region is field-free and consists of two grids placed at ground potential, an extraction grid and a second grid placed just prior to the detector.

In the time-of-flight mass analyzer, the velocity of an ion is used to determine the mass-to-charge ratio (m/z). As already stated above, a packet of ions is accelerated to constant kinetic energy by an electric potential, this gives the following formula from the law of conservation of energy:

Energy from Accelerating voltage = Kinetic Energy

$$\mathbf{Ve = \tfrac{1}{2} mv^2} \quad \textbf{(eqn. 1.1)}$$

$$\mathbf{v = (2Ve/m)^{1/2}} \quad \textbf{(eqn. 1.2)}$$

where V is the potential applied, e is the charge of an electron, m is the mass of the ion and v is the velocity of the ion. As the ions pass through the field-free region, they travel with a velocity proportional to $(m/z)^{-1/2}$, separating into a series of spatially discrete individual ion packets. Each ion packet then strikes the detector and produces a signal that is a function of time, yielding a TOF mass spectrum. The velocity of the ion can be calculated by the flight path distance (d) travelled divided by the elapsed time (t):

$$\mathbf{v = d/t = flight\ path\ length\ /\ flight\ time}\quad \textbf{(eqn. 1.3)}$$

Rearranging equation 1.2 above with the velocity relationship of equation 1.3 gives the time of flight of an ion with mass m.

$$\mathbf{d/t = (2Ve/m)^{1/2}}\quad \textbf{(eqn. 1.4)}$$

$$\mathbf{t = [d/(2Ve)^{1/2}]\times[(m)^{1/2}]}\quad \textbf{(eqn. 1.5)}$$

Simplifying eqn. 1.5 by substituting the constants with 'a' and accounting for multiply charged ions, the following formula relates the mass-to-charge ratio with the flight time for an ion:

$$\mathbf{t = a(m/z)^{1/2}}\quad \textbf{(eqn. 1.6)}$$

Ion flight times depend upon the square root of their masses. Lighter ions have higher velocities will arrive at the detector earlier than heavier ions carrying the same number of charges.

1.1.3 Mass Accuracy, Mass Calibration and Mass Resolution

In MALDI-TOF, the flight time can be affected by ion source conditions such as surface distortion (sample morphology) and laser power, thus the use of a proper mass scale for calibration is important for accuracy. The high mass accuracy of the TOF, in the ppm range, is due to two main factors: the high mass resolution and the simplicity of the mass calibration scale.[2,26,27] Mass accuracy is a measure of the error involved in the mass assignment of an ion signal. Often expressed as a percentage or part-per-million (ppm), it is the ratio of the error between the experimental and theoretical mass divided by the mass of the ion. The mass scale follows a square root law regardless of the relative sizes of the extraction and drift regions, or whether any other focusing elements are used (e.g., multiple-stage extraction, Einsel lenses, or reflectrons). Thus, mass spectra can be calibrated by measuring the flight times of two known masses to determine the constants a and b (where b accounts for time offsets due to laser firing time, triggering of recording devices, etc.) in the following formula:

$$\mathbf{t = am^{1/2} + b \quad (eqn.\ 1.7)}$$

In the linear mass spectrometer, the mass resolution obtained is governed by the width of an ion signal observed in the mass spectra. During MALDI, the ions formed in the source will be at different distances from the detector caused by a spatial spread of ions being formed (leading to different extraction time), different ion production time and have different imparted initial kinetic energies (resulting in initial velocity distribution). Ions with the same m/z value but with different distances to the detector will be detected at different times, resulting in peak broadening and decreased resolution observed by the detector.

Mass resolution, defined as $m/\Delta m$, is a measure of the instrument's ability to produce separate signals from ions of adjacent mass. It is expressed as the mass, m, of a given ion signal divided by the full width at half-maximum (FWHM) of the signal, Δm. In a TOF mass spectrometer in which ions are accelerated to constant kinetic energy, the following formula for resolution applies, where m and t are mass and flight time of the ion, Δm and Δt are the peak widths measured at the 50% level on the mass and time scales respectively:

$$\mathbf{m/\Delta m = t/2\Delta t} \quad \textbf{(eqn. 1.8)}$$

Other factors affecting resolution include the stability of power supplies, mechanical precision and the quality of the detector. In linear mode, the TOF is a low resolution mass analyzer. Many instruments incorporate multiple extraction regions as well as a number of focusing elements to improve mass resolution. Technological advances in MALDI TOF MS analysis indicate that the initial energy spread of the ions is the major factor limiting the resolving power of a TOF mass analyzer. Instruments employing reflectron and time-lag focusing devices have become valuable tools in reducing peak broadening caused by the initial velocity distribution of ions. The improved signal-to-noise ratio is due to grouping of ions into narrower peaks (increasing the peak height).

The reflectron or ion mirror was first introduced in 1973 by Mamyrin and coworkers[28] The ions are not detected after a single passage through the field-free drift region, as in the linear mode. Instead, the ions are reflected back into the field-free drift tube by the electric field of the reflectron and ion detection occurs near the start of the drift tube (refer to Figure 1.2). The reflectron consists of a series of parallel electrodes to which gradually increasing voltages is applied. For ions of the same m/z, more energetic ions move faster in the drift region, but penetrate deeper into the mirror, and as a result stay longer in it than less energetic ions. Ions with the same m/z value can then be focused at the detector. In this configuration, the reflectron extends the flight path which also improves mass resolution.

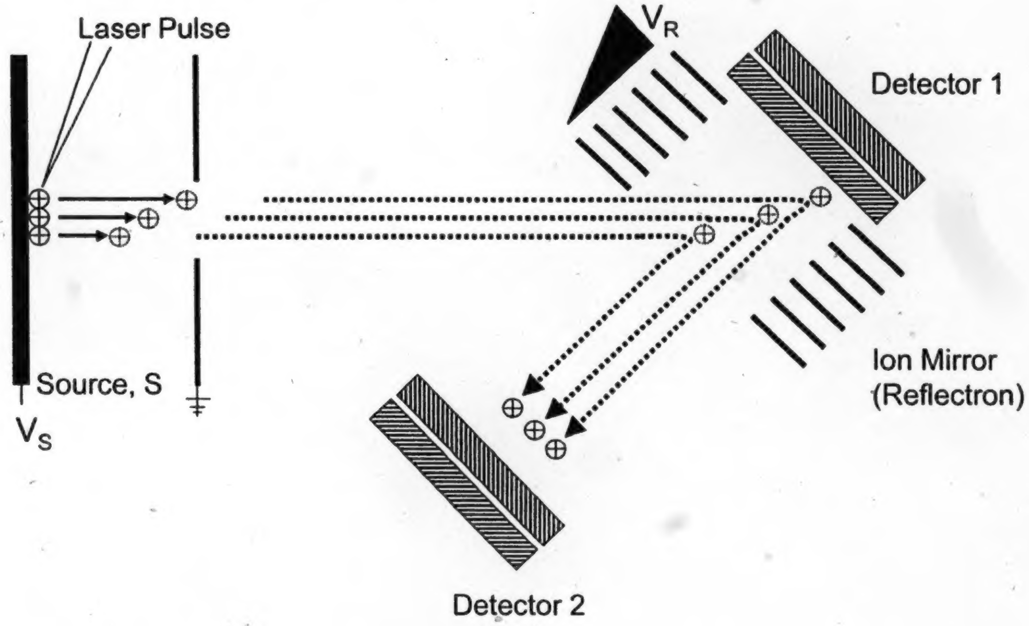


Figure 1.2 Schematic of reflectron mode.

In 1955, Wiley and McLaren[29] described an instrument with a two-stage accelerating region, which is now adopted in TOF systems to improve resolution. Several related designs of time-lag focusing have been applied to MALDI TOF.[30,31,32,33] In the conventional MALDI TOF MS configuration, the ions are immediately extracted from the source following ion production. This mode of operation, known as time-lag focusing or delayed extraction, establishes a delay between the end of the ionization pulse and the start of the extraction pulse so that the ions could be sorted in space, according to their initial kinetic energy. After a time delay, the extraction pulse imparts more energy to an ion that starts further from the field-free region, and will then be accelerated to a slightly higher energy than an ion with the same m/z value that is initially closer to the field-free region. By allowing a delay for the slower ions to catch up to the faster ones, ions with the same m/z will strike the detector at the same time despite differences in their initial kinetic energies.

TOF is considered to be a high-speed mass analyzer capable of high sensitivity. Limited by the flight time of the heaviest ions, usually in the 100 μs range, thousands of full mass spectra can be generated per second. A practical sensitivity of picomoles of protein can be readily achieved.[2] The high sensitivity of TOF mass analyzers, as compared to a scanning analyzer such as the quadrupole, is attributed to high ion transmission since all the m/z values in the flight tube can be detected. However, for very high m/z ions, the unlimited mass range offered by the TOF depends upon the quality of the ion detector and proper acceleration potential settings.

1.1.4 Detector

The detector used in the TOF and Q-TOF instruments is the microchannel plate (MCP) detector[34] consisting of two microchannel plates made of a lead-glass inner surface placed in a chevron configuration.[27] Each ion that strikes the MCP creates a pulse of electrons at the anode. The pulse is amplified and is used to trigger a timing pulse which is sent to the ion counting device, typically transient recorders (TRs) and time-to-digital converters (TDCs) are used to digitize the ion current to form a mass spectrum. To be detected, the large ions generated by MALDI must be converted into either electrons or low-mass ions at a conversion electrode. These electrons or low-mass ions are then used to start the electron multiplication cascade for amplification. The yield of secondary electrons from the conversion electrode is a function of the velocity of the ions to be detected.[35,36] To increase the detection sensitivity or increase the yield of secondary electron generation, an acceleration potential may be increased or by applying a post-acceleration in the flight tube. The conversion electrode should have a flat surface and is placed orthogonal to the ion trajectory, so that the ion flight pathlength is kept constant across the surface of the detector for accurate flight time recording. Different degrees of penetration of ions into microchannels may result, producing a significant time spread and decreasing the resolution.[27] Therefore, MCPs with the smallest channel diameter should be used.

1.2. Introduction to Quadrupole-TOF Mass Analyzer

Introduced in the early 1950s, the quadrupole mass analyzer is used as both a stand-alone device and in multistage mass spectrometers like the triple quadrupoles[37,38] and quadrupole TOF (Q-TOF) instruments[39,40]. Q-TOF mass spectrometers were first used for the analysis of peptides and now have a wide range of applications from biological samples in nanospray analysis to pharmaceutical preparations in liquid chromatography (LC)/MS/MS. Their popularity is due to the high sensitivity, mass accuracy and mass resolution of the TOF instruments for both precursor and product ions and also from their ability to record all ions in parallel, without scanning. Triple quadruple systems still offer the highest absolute sensitivity for targeted compounds (requiring the measurement of only a few types of ions), however, for certain analytical situations, the Q-TOF may provide better S/N due to the increased specificity afforded by the higher resolution of Q-TOF systems.[27]

1.2.1 Quadrupole Mass Analyzer

The quadrupole mass analyzer consists of a set of four electrodes of circular cross section positioned in a radial array, such that it approximates the ideal hyperbolic field. The quadrupole operates with both radio frequency (rf) and direct current (dc) voltage components applied to the rods (refer to Figure 1.3).

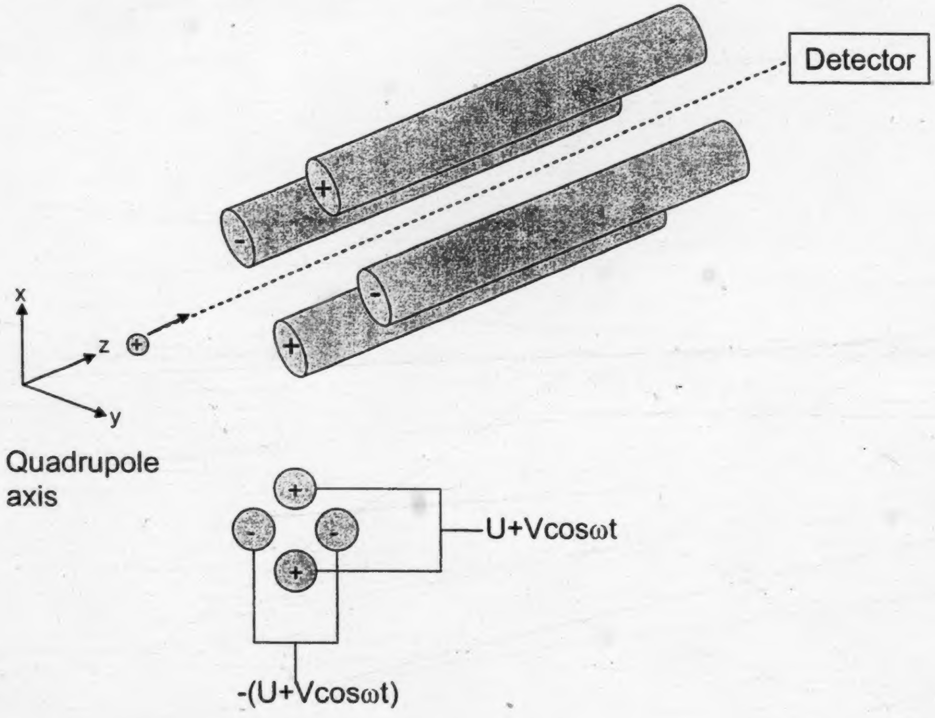


Figure 1.3 Schematic of quadrupole mass filter.

To one rod pair(+), a positive dc voltage and a time-dependent rf voltage are applied. To the other rod pair (-), a negative dc voltage and a time-dependent rf voltage are applied.

The ion experiences a force proportional to its distance from the center of the oscillating electric field. If the amplitude and frequency at which the field oscillates are sufficient, the ion will be accelerated away from one pair of electrodes towards the other pair. of oppositely charged electrodes and be confined between the electrodes. Thus, in the two-dimensional quadrupolar field, the ion will be confined along two axes, X and Y axes, while drifting along the third axis, Z axis.[41]

The nature of the potential distribution (Φ) at a given time (t) that ionized particles encounter in the quadrupole is described in the following formula, where U and V are the magnitude of dc and rf voltage respectively, ω is the angular frequency of the applied rf waveform, x and y are the distances along the given coordinate axes, r_o is the distance from the center axis (the z axis) to the surface of any electrode[41]:

$$\mathbf{\Phi = [U + V\cos(\omega t)](x^2 - y^2)/2r_o^2 \quad (eqn.\ 2.1)}$$

The theory of the quadrupole mass filter is governed by differential equations describing the trajectory of particles within the device. The generalized behavior of the solutions of such differential equations is represented by the reduced Mathieu parameters, a and q, describing the motion of an ion with mass (m/z) and charge (e).

$$\mathbf{a = 4eU/\omega^2 r_o^2 m \text{ and } q = 2eV/\omega^2 r_o^2 m \quad (eqn.\ 2.2)}$$

The solutions of this type of differential equation may be considered as being either bounded solutions or unbounded solutions. A bounded solution corresponds to a stable trajectory of particles as it is transmitted through the quadrupole for detection. An unbounded solution corresponds to particles which have unstable trajectories and will be deflected due to collisions with the electrodes, before they can strike the detector. Combinations of a and q that have stable solutions of the equations of motion in both X-Z and the Y-Z plane have been plotted in the stability diagram (refer to Figure 1.4), a plot of regions in a-q space where solutions to the equations of motion are stable or unstable.[42]

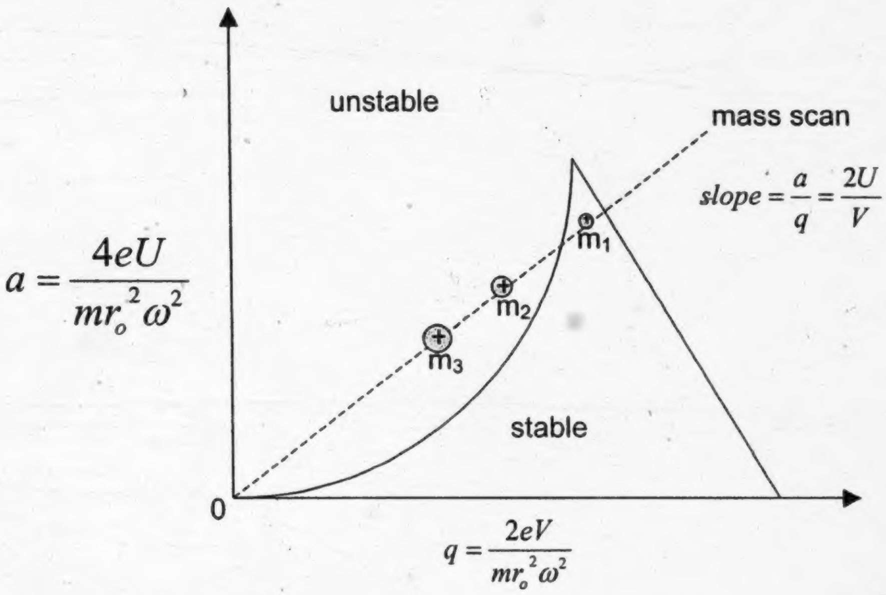

unstable
mass scan
$slope = \frac{a}{q} = \frac{2U}{V}$
$a = \frac{4eU}{mr_o^2 \omega^2}$
m_1
m_2
m_3
stable
0
$q = \frac{2eV}{mr_o^2 \omega^2}$

Figure 1.4 Plot of stability diagram.

The stability diagram allows the operation of the quadrupole mass filter to be simplified to two parameters, a and q, which is then further simplified by the two components of the potential applied, U and V. Quadrupoles are usually operated by holding the ratio U/V constant, which is equivalent to restricting operation of the mass filter to a set of operating points which lie on a straight line with a zero intercept. Such a line is known as the mass scan line. Simplifying common factors, the slope of the mass scan line (the quantity a/q) is reduced to the ratio 2U/V. If the slope is lowered, more ions covering a wider range of m/z pass through, and an increase in the signal is detected but the resolution is reduced.

The apex of the a-q diagram may be used to create a narrow bandpass mass filter. The bandpass region of this filter is changed by adjusting the slope of the mass scan line so that only particles within a certain m/z range will have stable trajectory to be detected. This is done by simultaneously increasing the applied voltages V and U, while keeping their ratio constant, then, the magnitude of the mass represented on the mass scan line also increases. From the kinetic energy relationship, the mass is inversely proportional to both a and q, the light ions are detected first followed by the heavier ones as the voltage is sweeped.[27]

When no dc potential is applied to the rods, the Mathieu parameter, a is equal to zero and the mass scan line is represented by a line with a slope of zero and intercepts the a axis at the point a=0. Thus, a large number of ions having different m/z values would be predicted to have stable trajectories within the rf-only quadrupole. Such a device is a high-mass filter, ions with m/z values below a certain cut-off value corresponding to q=0.908 are rejected. However, transmission of ions with high m/z values suffers because of poorer focusing[43] and can be lost in any of the three quadrupoles on their way to the TOF analyzer.

1.2.2 Basic Operation of MALDI QqTOF MS and MS/MS

The QqTOF instrument[27], named for its operation in MS/MS mode, consists of a mass-resolving quadrupole (Q), an r.f.-only quadrupole (q) and a reflecting time-of-flight

(TOF) mass spectrometer with orthogonal injection of ions[44,45]. There is an additional rf quadrupole Q0 is added to provide collisional damping and focusing of ions, so the instrument consists of three quadrupoles, Q0, Q1, and Q2. The ions are thermalized in collisions with neutral gas molecules, usually argon or nitrogen, reducing both the energy spread and the beam diameter and resulting in better transmission into and through both the quadrupole and TOF analyzers.

In a single-MS mode with Q0, Q1 and Q2 in the r.f.-only mode, the quadruples serves merely as a transmission element, while the TOF analyzer is used to record spectra. Ions covering approximately an order of magnitude in m/z range are transmitted simultaneously into the TOF instrument. If a wider m/z range is required, the rf-voltage is modulated (stepped or ramped between two or more rf levels) during the spectrum acquisition, providing a larger transmission window averaged over time. For MS/MS, Q1 is operated in the mass filter mode to transmit only the parent ion of interest. The ion is then accelerated to an energy of between 20 and 200 eV before it enters the collision cell Q2, where it undergoes collision-induced dissociation (CID) after the first few collisions with neutral gas molecules. The resulting fragment ions and any remaining parent ions are collisionally cooled and focused. This step is even more important in QqTOF instruments than it is in triple quadrupoles since the quality of the incoming ion beam is more crucial in the TOF analyzer (as discussed above) than it is in Q3 in a triple quadrupole instrument. An important advantage of QqTOF instruments over triple quadrupoles is the high mass resolution of TOF, typically around 10,000. After leaving the collision cell, ions are re-accelerated to the required energy (usually several tens of eV per unit charge), and focused by ion optics. The parallel ion beam continuously enters the ion modulator of the TOF analyzer, equipped with time-lag focusing and reflectron operations, and is recorded at the MCP detector. Since both MS and MS/MS spectra are recorded in the TOF analyzer, the same high resolution, mass accuracy, and mass calibration is achieved in both modes.[27]

1.2.3 Fragmentaton in MS/MS

Peptide mass mapping is often not sufficient for unambiguous identification of proteins and thus requires MS/MS for confirmation. The development of tandem mass spectrometry (MS/MS) is the coupling of two mass analyzers. The first MS device serves to isolate the target peptide from other peptides and matrix ions before the peptide (the parent ion) is accelerated through a region (the collision cell) where it undergoes collisions with a gas. The resulting fragment ions (the daughter ions) are analysed in the second MS device. The fragments ions can be subsequently used to determine the amino acid sequence of the precursor peptide which is then used to determine the identification of the protein. The daughter ions can be selected for further MS/MS analysis (MS^n) for a more detailed structure elucidation in some tandem MS analyzers such as ion trap.

Different types of fragment ions and thus different sequence information is obtained whether low or high energy is used for fragmentation. We will restrict our discussion of fragmentation to low-energy regimes, since this mode is more common in practice. The set of fragment ions produced by low-energy collisionally induced dissociation (CID) are shown in Figure 1.5. The nomenclature shown were first classified by Roepstorff and Fohlman[46] (1984) and later modified by Biemann[47] (1990).

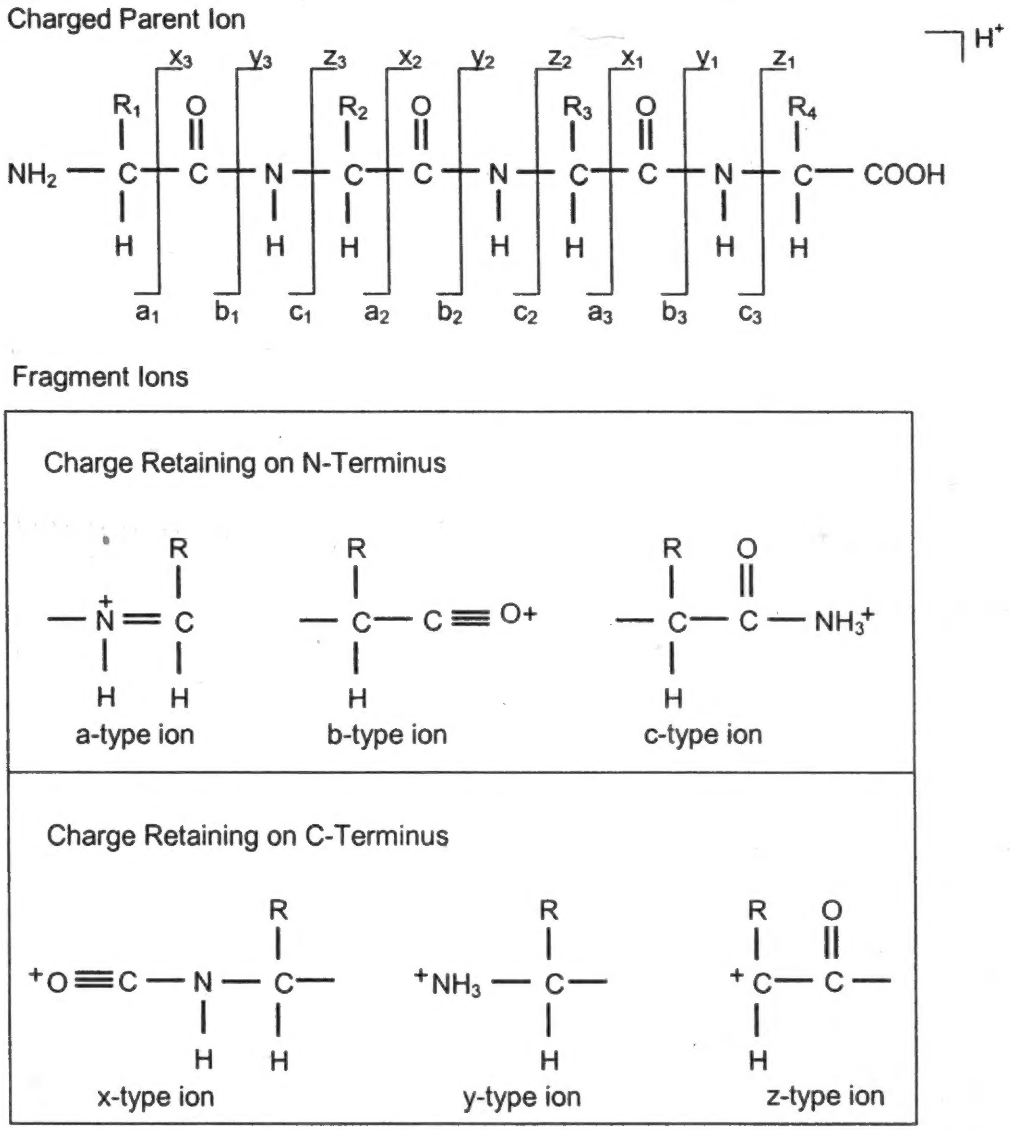

Charged Parent Ion
H+
x3 y3 z3 x2 y2 z2 x1 y1 z1
R1 R2 R3 R4
NH2 C C N C C N C C N C COOH
a1 b1 c1 a2 b2 c2 a3 b3 c3
Fragment Ions
Charge Retaining on N-Terminus
a-type ion
b-type ion
c-type ion
Charge Retaining on C-Terminus
x-type ion
y-type ion
z-type ion

Figure 1.5 Low energy CID fragmentation nomenclature.

When the peptide ions collide with the gas in the collision cell, they undergo multiple (one to ten) collisions and the kinetic energy is converted into vibrational energy. The ions become excited to an unstable state and the energy is then sufficient to cleave one of the amide bonds in the peptide backbone. The subsequent fragmentation pattern depends on the collisonal energy, the pressure and type of collision gas, the number of charges being carried by the peptide, and the amino acid sequence of the peptide.[48] For singly charged peptide ions, the daughter products formed consists of an ion and a small neutral molecule. Only those fragments carrying a charge are detected in the spectrum. If the charge is retained on the N-terminus, the fragment ions are classified as either a, b, or c-type ions. While fragment ions of x, y, or z-type ions have the charge retained on the C-terminus. Low energy CID mainly generate b and y-type ions. If the collision energy is relatively high, then fragmentation involving a single bond cleavage is favoured, and the preferential formation of b-type ions occurs. However, if the energy is relatively small, a pathway with a lower activation energy is favoured, and y-type ions are more likely to form.

In the gas-phase, the nitrogen atom of the amide moiety of the peptide backbone and of the primary amine have equivalent proton affinity. There is free movement of protons from the N-terminus and Lys side chains, resulting in a random proton distribution along the backbone. The distribution of charge is important in determining the relative amounts of b- and y-type ions that are detected. Also a factor is the number and position of residues which have a high basicity in the gas phase, such as Pro, His, Trp, Arg and Lys. In the low-energy regime in a triple quadrupole mass spectrometer, the neutral and charged products do not separate immediately after bond cleavage. The close proximity of the products to each other result from dipole-charge attractions until further collisions cause separation.[48] During this transition state, proton transfer to the residue with the highest basicity can occur, which may explain the relative dominance of y-type ions in tryptic peptides that usually have Arg or Lys as the C-terminal residue.

The generation of fragment ions is enhanced by the site of protonation on a peptide backbone because of better ionization efficiency and ion stability at these sites.[49a] In contrast, b-type ions are more dominate in a spectrum of a peptide obtained using a quadrupole ion trap mass spectrometer.

In addition to the six sequence-specific ions, a number of side-chain cleavages can also occur, forming the N-terminal d ions and the C-terminal v and w ions, but these only occur in high-energy regimes.[48] Ions formed by internal cleavages are those that result from multi-point cleavage of the backbone, which include internal acyl ions ($R-CO^+$) and internal immonium ions ($R-^+N=C$), including single amino acid immonium ions. These immonium ions arise from at least two internal bond cleavages and occur in the low-mass region (m/z <200) of MS/MS spectra. They are diagnostic of the presence of specific amino acids in the peptide and important for interpretation. In particular, the immonium ions of His (m/z 110) and Pro (m/z 70) are usually quite intense.

Ions termed b^o and y^o (formed by the loss of water) and b* and y* (if ammonia is lost) often occur in low-energy spectra, but rarely occur in high-energy spectra. The preferential cleavage C-terminal to Asp is often observed. In the gas phase, the carboxylic acid group of Asp is often involved in rearrangement reactions with the amide bond, giving rise to a cyclic intermediate that decomposes to yield a b-type ion that often dominates the MS/MS spectrum. For tryptic peptides, the proton is localized on Arg and Lys residues and y-type ions predominate the spectra. The preferential cleavage of singly charged tryptic peptide ions C-terminal to aspartate residues was reported.[49b] The opposite is true for Pro, where almost no C-terminal cleavage is found. Similarly, no C-terminal cleavage of R or K residues next to Pro is observed in trypsin digestion. In addition, the specific loss of certain groups may indicate the presence of certain modifications, such as phosphorylation and glycosylation.

1.3 Separation Methods

Most single-dimension separations exhibit poor resolution for analysis involving complex cellular extracts, and therefore multi-dimension separations must be employed. Dynamic range and protein solubility issues complicate the separation of low-abundance and hydrophobic proteins and their detection by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).[50,51,52] The analysis of integral membrane proteins remains a challenge. New detergents have been designed to enhance membrane protein solubility for analysis by 2D-PAGE.[53,54,55] The use of narrow pH-range IPG gels and sample loading via rehydration allows for higher amounts of protein to be loaded and thus extends the sensitivity range. Another approach to solving this problem is to prefractionate complex protein mixtures before loading on 2D gels, thus increasing the amounts of low-abundance proteins relative to other proteins, but this is still quite labour intensive.

Alternative separation methods to gel electrophoresis, particularly, reversed phased liquid chromatography (RP LC), offer complementary information. A different selection of proteins will be identified to that derived from 2D PAGE. The choice of separation method to employ depends on the analytical task and on the biological system. Despite development of new methods, the 'traditional' methods still offer analytical advantages. Some newer methods include the use of combinations of chromatographic or electrophoretic separations in the liquid phase, such as two dimensional liquid chromatography and hybrid electrophoretic-chromatographic approach (IEF and NP-RP-HPLC).[56] The emergence of multidimensional liquid chromatography (LC) provide direct interface of protein and peptide separations to mass spectrometers.[57] However, these methods yield lower resolution of proteins from complex mixtures, and provide inferior visualization of the total protein expression within a biological system, as compared to 2D PAGE.

1.3.1 2D Polyacrylamide Gel Electrophoresis

2D-gel electrophoresis was invented in 1975 by O-Farrell.[58] 2D-PAGE is widely used for the global study of protein expression in biological systems. This method usually involves enzymatic digestion of the excised protein spots, peptide mass mapping[59] and/or tandem mass spectrometry[60] followed by bioinformatics searches for unambiguous protein identification. The quality of 2D separations is highly dependent on good sample preparation. The advantage is that thousands of proteins and their isoforms (including post-translational variants) from a cell or tissue can be resolved in a gel. Cells exhibit a large dynamic range of expression, a single cell may express 10,000 or more proteins at the same time. However, the limitation of this method is that it can not monitor all types of proteins in the biological system at the same time in a gel. Proteins not readily analyzed by this strategy include membrane, low copy number, highly basic, and very large (>150 kDa) and small (<10 kDa) proteins.[56]

1.3.2 Basic Principles of Isoelectric Focusing

The focusing effect of IEF, which concentrates the proteins at their pIs, allows for proteins to be separated on the basis of very small charge differences. The protein mixture is loaded onto an immobilized pH gradient (IPG), typically 3-10 pH range. IPGs were introduced by Bjellqist and coworkers[61] in 1982, and are created by covalently incorporating a gradient of acidic and basic buffering groups to an acrylamide monomer. During polymerization, the acrylamide portion of the buffers copolymerize with the acrylamide and bisacrylamide to form a polyacrylamide gel which is cast onto a plastic backing and frozen until use. The IPG strip is rehydrated in a rehydration solution containing sample prior to performing IEF. An electric field is applied and the proteins will move to their respective positions in the gradient, at the pH at which their net charge is zero, or at its pI. It is during the focusing step of IEF that membrane proteins are susceptible to precipitation.[62] After IEF, the IPG strips with focused proteins are equilibrated in an equilibration solution and ready for SDS-PAGE. IEF performed under denaturing conditions provides clean results with high resolution. Denaturation is achieved with a mixture of urea, a non-ionic detergent and a reductant added to the rehydration solution to ensure that each protein is present in only one configuration. This minimizes aggregation and intermolecular interactions that would result in streaking of the gel.[63]

1.3.3 Basic Principles of SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)[64] is routinely used for the estimation of protein molecular weights and for purifying of protein mixtures into separate components. Polyacrylamide gels are formed by copolymerization of acrylamide monomer, CH_2=CH-CO-NH_2, and a cross-linking bifunctional comonomer, N,N'-ethylene-bis-acrylamide (Bis), CH_2=CH-CO-NH-CH_2-NH-CO-CH=CH_2. The mechanism for polymerization is vinyl addition catalyzed by a free radical generating system composed of an initiator, ammonium persulfate (APS) and an accelerator, tetramethylethylenediamine (TEMED). TEMED causes the formation of free radicals from persulfate and these in turn catalyze polymerization. Oxygen is a radical scavenger and interferes with polymerization, so that proper degassing to remove dissolved oxygen from acrylamide solutions is crucial for reproducible gel formation. The sieving properties are established by the pores of a gel. Within limits, holding the cross linker concentration (%C) constant, the pore size of a gel decreases as the amount of total acrylamide monomer concentration (%T) increases.[65]

In an electrophoretic separation, charged particles migrate toward the electrode of opposite sign under the influence of an externally applied electric field. The movements of the particles are retarded by interactions with the surrounding gel matrix, and the opposing interactions cause differential migration rates of proteins in a sample mixture. The gel serves as a supporting medium that minimizes convection currents and diffusion to warrant the separated components remain as highly resolved, sharp zones. Ideally, the gel should be chemically inert during the separation process. However, formation of protein-acrylamide adducts from the covalent bonding of residual acrylamide monomer to cysteine residues on proteins have been observed, complicating succeeding MS analysis.[66]

In general, fractionation by gel electrophoresis is based on the sizes, shapes, and net charges of the macromolecules. Uniform hydrodynamic and charge characteristics on all the proteins in a sample mixture is imposed by the denaturing treatment of SDS and heat. The anionic detergent binds tightly to most proteins at ~1.4 mg of SDS to 1 mg of protein and imparts a negative charge to the resultant protein-SDS complex. Interaction with SDS disrupts all non-covalent protein bonds, causing the macromolecules to unfold. Treatment with a disulfide-reducing agent, such as 2-mercaptoethanol or dithiothreitol, further denatures proteins by breaking any disulfide bonds. The electrophoretic mobilities of the resultant detergent-polypeptide complexes depend largely on their molecular weights. Migration of anionic SDS-protein complexes is toward the anode (positive electrode) at rates inversely proportional to the logarithms of their molecular weights (over a limited range), with low molecular weight complexes migrating faster than larger ones. The molecular weight of a protein can be estimated from its relative mobility (distance of the protein band) in a calibrated SDS-PAGE gel. The most common buffer system used is the Tris-glycine system described by Laemmli[67], which is the system used in our experiments.

Once excised, the gel slices are washed with a solution of ammonium bicarbonate and acetonitrile, in order to remove excessive Coomassie Blue dye and to bring up to proper pH for trypsin digestion.[68,69] The cleavage fragments are extracted out of the gel slice with acidic and organic solvents by passive diffusion. A basic extraction using 0.1 M ammonia was reported to be slightly less effective.[70] An enzymatic fragmentation generates a peptide mixture characteristic of the parent protein; it is called a "peptide map", a "fingerprint" or a "mass map" of a protein. The recovery of peptide fragments from a polyacrylamide gel by solvent elution is based on solubility and hydrophobicity of peptides. The same is true for separations by RP-HPLC as discussed in the subsequent section. The recovery of the resulting peptides is achieved by elution from the gel matrix with organic solvents and separation by RP-HPLC or can also be directly analyzed using MALDI-MS. The HPLC chromatogram, the gel pattern or a mass spectrum of the mixture can be used for comparative studies.

1.3.4 RP HPLC

Separations in reversed-phase liquid chromatography are based on polarity. The stationary phase interacts strongly with the sample molecules depending on the polarity of the sample and elution solvent. The bonded phase is non-polar, (commonly, C8 or C18 silica columns) and the eluents are usually polar, often a combination of water and acetonitrile. On these columns, the non-polar sample molecules prefer the stationary phase; the lower the sample polarity, the greater the retention time. To separate a protein mixture containing many components of various polarities, eluent polarity is changed during the experiment. To bring a non-polar sample through the column faster, a less polar eluent is used. Trifluoroacetic acid (TFA) is often used for ion-pairing of peptides and proteins, resulting in improved peak shapes and higher resolution.

Reversed-phase liquid chromatography separations can easily be interfaced on-line to electrospray, or off-line to MALDI sources because of compatibility of solvents and lack of salts and detergents. The advantages of using liquid chromatography for protein separation are high sample loadability, speed, reproducibility and ease of automation. The improved spectral quality from the LC separation step reduces ion suppression in an LC-MALDI MS experiment. Some proteins ionize more easily than others and subsequently suppress the signal from the less ionizable proteins and complicate their identification. However, difficulty in membrane protein analysis and in resolving proteins from entire cellular systems are inevitable in chromatographic methods.[56]

1.4. Scope of the Thesis Work

In this work, proteomic analysis was applied to membrane proteins and cancer cell lines. The application of gel electrophoresis or liquid chromatography in combination with mass spectrometry towards the characterization of these biological systems remains a valuable technique in proteome research due to its simplicity and easy sample handling. Membrane proteins play a key role in cell adhesion, signal transduction, and molecular transportation across the lipid bilayer. Generally, membrane proteins are hydrophobic and have an alkaline pH. This makes their analysis difficult because they are not soluble in polar solvents, requiring detergents to help stabilize and solubilize hydrophobic proteins in solution and prevent aggregation. Their low abundance in cells and low copy numbers of individual proteins further complicates membrane protein analysis.

In Chapter 2, a sequential CNBr and trypsin in-gel digestion method combined with mass spectrometry is presented for membrane protein analysis. To increase the amount of peptides detected, it was demonstrated that trypsin can be used to further digest the sample in gel after CNBr cleavage. In addition, the use of *n*-octyl glucoside (*n*-OG) to enhance the digestion efficiency and peptide recovery was also studied. We demonstrate that the sensitivity of this membrane protein identification method is in the tens of picomole regime which is compatible to Coomassie staining gel-spot visualization method, and is more sensitive than other techniques reported in the literature. This CNBr/trypsin in-gel digestion method is also found to be very reproducible and has been successfully applied for the analysis of complex protein mixtures extracted from biological samples. The results are presented from a study of the analysis of bacteriorhodopsin (BR), nitrate reductase 1 gamma chain (NAR I *E. coli*), and a complex protein mixture extracted from the endoplasmic recticulum (ER) membrane of mouse liver.

The profiling of proteomes provides important information in studies of protein expression and diseased tissue. To reduce complexity of whole cell lysates, fractionation allows the visualization of low abundance proteins. In Chapter 3, the profiling of changes in protein expression of SCC9 cells transfected with cell adhesion proteins, plakoglobin (Pg) and E-cadherin (E-cad), since these proteins were lacking in the transformed cells, was undertaken. The study is used to obtain a better understanding of the cadherin-catenin system in the regulation of cell proliferation, invasion, and intracellular signalling during cancer. Three different approaches were performed on the squamous carcinoma cells to investigate and compare the differences in the proteins observed for SCC9, SCC9-E-cad, and SCC9-Pg cells. In this work, the use of 2D-gel electrophoresis generated a proteomic map of the extracted proteins in the 20-100 kDa mass range while the methods of direct MALDI MS and RP HPLC offline MALDI MS generated proteins in the mass range below 20 kDa. The proteomic profiling of these transformed cells has not been previously investigated in the literature.

1.5 Literature Cited

(1) Vastola, F. J.; Mumma, R. O.; Pirone, A. J. *J. Org. Mass Spectrom.* **1970**, *3*, 101-104.

(2) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A-1203A.

(3) Beckey, H. D. *Principles of Field Desorption Mass Spectrometry*; Pergamon Press: Oxford, 1977.

(4) Macfarlane, R. D.; Torgerson, D. F. *Science* **1976**, *191*, 920.

(5) (a) Benninghoven, A.; Sichtermann, W. K. *Anal. Chem.* **1978**, *50*, 1180. (b) Barber, M.; Bordoli, R. S.; Sedgwick, R. S.; Tyler, A. N. *J. Chem. Soc. Chem. Commun.* **1981**, 325.

(6) Blakely, C. R.; Carmody, J. J.; Vestal, M. L. *Anal. Chem.* **1980**, *52*, 1636.

(7) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal. Chem.* **1985**, *57*, 675.

(8) Cotter, R. J. *Anal. Chim. Acta* **1987**, *195*, 45-49.

(9) Karas, M.; Bachmann D., Bahr, U.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53-68.

(10) Tanaka, K.; Ido, Y.; Akita, S. In *Proceedings of the Second Japan-China Joint Symposium on Mass Spectrometry*; Matsuda, H.; Liang, X.-T., Eds.; Bando Press: Osaka, Japan, 1987: pp 185-188.

(11) Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.

(12) Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299-2301.

(13) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science*, **1989**, *246*, 64-71.

(14) Beavis, R. C.; Chaudhary, T.; Chait, B. T. *Org. Mass Spectrom.* **1992**, *27*, 156-158.

(15) Beavis, R. C.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1989**, *3*, 432-435.

(16) Strupat, K.; Karas, M.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1991**, *111*, 89.

(17) Weinberger, S. R.; Boernsen, K. O.; Finchy, J. W.; Robertson, V.; Musselman, B. D. In *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*; San Francisco, CA, May 31-June 4, 1993; pp 775a-b.

(18) Xiang, F.; Beavis, R. C. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 199-204.

(19) Xiang, F.; Beavis, R. C. *Org. Mass Spectrom.* **1993**, *28*, 1424-1429.

(20 Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, *66*, 3281-3287.

(21) Edmondson, R. D.; Campo, K. K.; Russell, D. H. In *Proceedings of the 41st ASMS Conference on Mass Spectrometry and Allied Topics*; Atlanta, Georgia, May 21-May 26, 1995; p 1246.

(22) Li, L.; Golding, R. E.; Whittal, R. M. *J. Am. Chem. Soc.* **1996**, *118*, 11662-11663.

(23) Kussmann, M.; Nordhoff, E.; Rahbek-Nielsen, H.; Haebel, S.; Rossel-Larsen, M.; Jakobsen, L.; Gobom, J.; Mirgorodskays, E.; Kroll-Kristensen, A.; Palm, L.; Roepstorff, P. *J. Mass Spectrom.* **1997**, *32*, 593-601.

(24) Dai, Y. Q.; Whittal, R. M.; Li. L. *Anal. Chem.* **1996**, *68*, 2721-2725.

(25) Zhang, N.; Doucette, A.; Li, L. *Anal. Chem.* **2001**, *73*, 2968-2975.

(26) Cotter, R. J. *Time-of-Flight Mass Spectrometry: Instrumentation and Applications in Biological Research*; American Chemical Society: Washington, D.C., 1997.

(27) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J. Mass Spectrom.* **2001**, *36*, 849-865.

(28) Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. *Sov. Phys. JETP* **1973**, *37*, 45-48.

(29) Wiley, W. C.; McLaren, I. H. *Rev. Sci. Instrum.* **1955**, *26*, 1150-1157.

(30) Colby, S. M.; King, T. B.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 865-868.

(31) Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044-1050.

(32) Brown, R. S.; Lennon, J. J. *Anal. Chem.* **1995**, *67*, 1998-2003.

(33) Whittal, R. M.; Russon, L. M.; Weinberger, S. R.; Li, L. *Anal. Chem.* **1997**, *69*, 2147-2153.

(34) Wurz, P.; Gubler, L. *Rev. Sci. Instrum.* **1996**, *67*, 1790-1793.

(35) Beuhler, R. J. *J. Appl. Phys.* **1983**, *54*, 4118-4126.

(36) Spengler, B.; Kirsch, D.; Kaufmann, R.; Karas, M.; Hillenkamp, F.; Giessmann, U. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 301-305.

(37) Yost, R. A.; Enke, C. G. *J. Am. Chem. Soc.* **1978**, *100*, 2274-2275.

(38) Yost, R. A.; Boyd, R. K. *Methods Enzymol.* **1990**, *193*, 154.

(39) Morris, H. R.; Paxton, T.; Dell, A.; Langhorn, B.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889-896.

(40) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015-1024.

(41) Miller, P. E.; Bonner Denton, M. *J. Chem. Educ.* **1986**, *63*, 617-622.

(42) McLachlan, D. W. *Theory and Applications of Mathieu Functions*; Oxford, 1947.

(43) Gerlich, D. *Adv. Chem. Phys.* **1992**, *82*, 1.

(44) Krutchinsky, A. N.; Loboda, A. V.; Spicer, V. L.; Dworschak, R.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 508.

(45) Loboda, A. V.; Krutchinsky, A. N.; Bromirski, M.; Ens, W.; Standing, K. G. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1047.

(46) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601.

(47) Biemann, K. *Methods Enzymol.* **1990**, *193*, 886-887.

(48) Kinter, M.; Sheerman, N. E. In *Protein Sequencing and Identification Using Tandem Mass Spectrometry*; Wiley-Interscience: New York, 2000.

(49) Sullivan, A. G.; Brancia, F. L.; Tyldesley, R.; Bateman, R.; Sidhu, K.; Hubbard, S. J.; Oliver, S. G., Gaskell, S. J. *Int. J. of Mass. Spectrom.* **2001**, *210/211*, 665-676 and (a) references therein {5-13, 21} and (b) references therein {14,17}.

(50) Gygi, A. P.; Rochon, Y.; Franza, B. R.; Aebersold, R. *Mol. Cell. Biol.* **1999**, *19*, 1720-1730.

(51) Santoni, V.; Molloy, M.; Rabilloud, T. *Electrophoresis* **2000**, *21*, 1054-1070.

(52) Gygi, S. P.; Corthals, G. L.; Zhang, Y.; Rochon, Y.; Aebersold, R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9390-9395.

(53) Rabilloud, T.; Adessi, C.; Giroudel, A.; Lunardi, J. *Electrophoresis* **1997**, *18*,

307-316.

(54) Chevallet, M.; Santoni, V.; Poinas, A.; Rouquie, D.; Fuchs, A.; Kieffer, S.; Rossignol, M.; Lunardi, J.; Garin, J.; Rabilloud, T. *Electrophoresis* **1998**, *19*, 1901-1909.

(55) Molloy, M. P.; Herbert, B. R.; Walsh, B. J.; Tyler, M. I.; Traini, M.; Sanchez, J.-C.; Hochstrasser, D. F.; Williams, K. L.; Gooley, A. A. *Electrophoresis* **1998**, *19*, 837-844.

(56) Nilsson, C. L.; Davidsson, P. *Mass Spectrom. Rev.* **2000**, *19*, 390-397 and references therein.

(57) Wolters, D. A.; Washburn, M. P.; Yates, J. R. III *Anal. Chem.* **2001**, *73*, 5683-5690.

(58) O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007-4021.

(59) Patterson, S. D.; Aebersold, R. *Electrophoresis* **1995**, *16*, 1791-1814.

(60) Courchesne, P. L., Patterson, S. D. *Methods Mol. Biol.* **1999**, *112*, 487-512.

(61) Bjellqvist, B.; Ek, K.; Righetti, P. G.; Gianazza, E.; Gorg, A.; Westermeier, R.; Postel, W. *J. Biochem. Biophys. Methods* **1982**, *6*, 317-339.

(62) Patton, W. F. *J. Chromatogr. B* **1999**, *722*, 203-223.

(63) Berkelman, T.; Stenstedt, T. In *2-D Electrophoresis: Principles and Methods*; Amersham Biosciences Handbook: Piscataway, NJ, 2002.

(64) Garfin, D. E. "One-Dimensional Gel Electrophoresis" In *Guide to Protein Purification: Methods in Enzymology*; Academic Press, Inc., 1990.

(65) Rodbard, C.; Chrambach, A. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *65*, 970.

(66) Clauser, K. R.; Hall, S. C.; Smith, D. M.; Webb, J. W.; Andrews, L. E.; Tran, H.

M.; Epstein, L. B.; Burlingame, A. L. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5072.

(67) Laemmli, U. K. *Nature* **1970**, *227*, 680-685.

(68) Rosenfeld, J.; Capdevielle, J.; Guillemot, J. C.; Ferrara, P. *Anal. Biochem.* **1992**, *203*, 173-179.

(69) Eckerskorn, C.; Lottspeich, F. Chromatographia 1989, 28, 92-94.

(70) Tetaz, T.; Bozas, E.; Kanellos, J.; Walker, I.; Smith, I. In *Techniques in Protein Chemistry IV*; (Angeletti, R., ed.); Academic Press: San Diego 1993; pp 389-397.

Chapter 2

Development and Applications of In-Gel CNBr/Tryptic Digestion Combined with Mass Spectrometry for the Analysis of Membrane Proteins[a]

2.1 Introduction

Material transport and signal transduction among cells occur via integral membrane proteins. These membrane proteins can provide key information that is of great interest to the researchers studying the functions of a biological system. Many mass spectrometric methods have been developed for protein identification, but they often fail to identify very hydrophobic membrane proteins. The challenge in the analysis of hydrophobic membrane proteins develops from their amphiphilic nature. They are not readily soluble in the polar solvents normally used in the digestion buffer and often undergo irreversible aggregation. As a result, loss of sample during the analysis is inevitable by protein adhesion to the sample handling surfaces. In addition, production and recovery of as many fragments as possible becomes more difficult for hydrophobic proteins due to poor accessibility for proteolytic attack and difficulties associated with the separation and detection of the resulting hydrophobic peptides.[1] Identification of some membrane proteins is still possible since a few soluble peptides generated from a membrane protein can be observed in the mass spectra. However, high sequence coverage is desired for reliable protein identification and good peptide mass mapping.

Both gel-based and solution-based methods have been used to analyze the membrane proteome with each having its own strengths.[2-5] It appears that the two methods are complementary to each other.[6] The commonly used in-gel approach of protein identification involves the separation of a protein mixture from cell or tissue extracts by one- or two-dimensional polyacrylamide gel electrophoresis (PAGE). The protein bands or spots are stained and excised for digestion. The resultant peptides are extracted from the gel piece and analyzed by MS and/or MS/MS. The masses or fragment ion spectra of these peptides are entered in a database search for protein identification.[7] This in-gel approach has some attractive features in analyzing membrane proteins. Strong surfactants can be used in sample preparation and gel separation, which can alleviate the problems of sample loss and irreversible aggregation of hydrophobic proteins. In addition, the proteins embedded within the gel matrix after separation establish a localized region for enzymatic or chemical cleavage and thus improve the chance for digestion. The gel acts as a molecular sieve in which the protein-SDS complex is embedded within the pores once electrophoresis is stopped and the protein is "fixed". When a same volume of digestion solution is used for in-gel and in-solution digestion, the protein concentration in a localized region of the gel band is higher for in-gel digestion than that of the homogenous protein solution for in-solution digestion. Protein unfolding and less interference after gel separation and subsequent washing are also the contributing factors which can lead to better results for some proteins by using the in-gel method.

---

[a] A form of this chapter has been published as Thuy Tien T. Quach, Nan Li, Dawn P. Richards, Jing Zheng, Bernd O. Keller and Liang Li "Development and Applications of In-Gel CNBr/Tryptic Digestion Combined with Mass Spectrometry for the Analysis of Membrane Proteins" *Journal of Proteome Research*, 2003, 2, 543-552. The preparation of the ER sample and the 1D gel separation was performed by Dr. Barbara Knoblach.

Although the in-gel digestion/MS approach for identifying hydrophilic proteins often uses 2D-PAGE for high-resolution protein separation, the separation of integral membrane proteins by 2D-PAGE is difficult and challenging. Severe loss of hydrophobic proteins is believed to be caused by protein precipitation at their isoelectric points during the focusing step of isoelectric focusing (IEF) and thus deters further migration in the second dimension.[8] As reviewed by Santoni *et al.* (2000), proteomics studies using 2D-PAGE of yeast, tobacco or Arabidopsis, transmembrane proteins could not be identified by MS or MS/MS. Only a few membrane proteins from mammalian cells have been successfully separated by 2D-PAGE and identified.[9] SDS-PAGE is therefore a better choice for handling very hydrophobic membrane proteins, albeit with reduced resolving power compared to 2D-PAGE. However, using organelle separation methods such as sucrose gradient technique, complicated membrane proteome can be fractionated for subsequent SDS-PAGE separation, thereby reducing the bundle on PAGE separation.

The real challenge in the in-gel digestion/MS approach for membrane protein identification lies in the process of in-gel digestion and subsequent sample preparation for MS analysis. We report herein an in-gel digestion protocol that involves the use of CNBr digestion, followed by further trypsin digestion. Due to its availability and hydrophobicity, Bacteriorhodopsin (~26.7 kDa), a light-driven proton pump with 7-helical-transmembrane regions[10], serves as a good model for method development and comparison. In the past, work has been performed on bacteriorhodopsin in-solution with a CNBr digestion, fractionation by HPLC and analysis by ESI-MS.[11] These authors show that complete sequencing of BR can be done with an in-solution CNBr-only digested sample of 5-10 nmole BR. However, this method is not adequate for the detection of the low abundance proteins, requiring at least nmol amount of pure starting material, due to severe sample loss from the many washing steps involved.[1]

A search of the literature suggests that CNBr followed by trypsin digestion of proteins has been performed only in-solution and only on hydrophilic proteins. Stone and coworkers in 1992 performed internal sequencing of PVDF-blotted proteins after using the combined chemical/enzyme digestion.[12] CNBr/trypsin digestion along with liquid chromatography/ESI mass spectrometry was performed to identify the active site serine of penicillin-binding protein 2a.[13] The CNBr digestion was performed in 0.1 M HCl for 2 h at 37 °C in the dark, the CNBr fragments were isolated by HPLC and further treated with trypsin in 0.1 M NH_4OAc (pH 8.5) for 30 min at 37 °C. The tryptic peptides were then analyzed by both microbore liquid chromatography/MS and nano-electrospray MS/MS. Another application of CNBr/trypsin digestion involved the separation of total cell extracts by SDS-PAGE, transfer of the separated proteins onto nitrocellulose membrane followed by excision of the protein band of interest for digestion with CNBr in 70% formic acid. The resulting CNBr-fragmented peptides were then fractionated by C18 reversed phase HPLC, followed by trypsin digestion and analysis by MALDI-TOF mass spectrometry.[14] In a related work, van Montfort *et al.* have reported an in-gel digestion method using trypsin followed by CNBr digestion.[15] Therefore, our work is the first demonstration of using sequential CNBr/trypsin digestion for the analysis of hydrophobic membrane proteins in-gel. This method is optimized to detect low amounts of membrane proteins that can be visualized in gel by Coomassie staining method. Detailed protocols and data interpretation are given in the examples shown herein, in the hope that others can readily adapt the method to analyze other membrane proteins of interest.

2.2 Experimental

2.2.1 Materials and Reagents

Sodium dodecyl sulfate (SDS) and α-Cyano-4-hydroxycinnamic acid (HCCA) were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). HCCA was recrystallized from ethanol (95%) at 50 °C prior to use. Analytical grade acetone, methanol, acetonitrile, and trifluoroacetic acid (TFA) were from Caledon Laboratories (Edmonton, AB). Water used in all experiments was from a NANOpure water system (Barnstead/Thermolyne). Bacteriorhodopsin (from Halobacterium Halobium), bovine trypsin, cyanogen bromide, glycerol, 2-mercaptoethanol, and *n*-octyl-β-D-glucopyranoside (*n*-OG) were obtained from Sigma (St. Louis, MO). Tris(hydroxymethyl)-aminomethane, acrylamide/bis (electrophoresis purity grade), and

Coomassie Blue G250 staining solution were obtained from BioRad (Hercules, CA). The peptide extracts were desalted using commercial C18 μZipTips from Millipore (Bedford, MA).

2.2.2 Bacteriorhodopsin

Bacteriorhodopsin (BR) was suspended in deionized water and the stock solution of 1 μg/μL was stored at –20 °C until use. The BR solution was denatured 1:5 with SDS sample buffer containing 2% (v/v) 2-mercaptoethanol, 1% SDS, 12% glycerol, 50 mM Tris-HCl (pH 8.8) and trace amounts of bromophenol blue and heated to 95 °C for 5 minutes before loading onto the gel. Gel electrophoresis was performed with SDS-PAGE (1 mm thick) mini-gels consisting of a 4% stacking gel and 12% separating gel and Coomassie blue stained. The excised gel piece was digested with trypsin, CNBr, trypsin/CNBr or CNBr/trypsin, as indicated in the Results and Discussion, and analyzed by MALDI MS or MS/MS as described below.

2.2.3 NAR I *E. coli*

Nitrate reductase 1 gamma chain (NAR I *E. coli*) was a gift of Professor Joel Weiner of the University of Alberta. NAR I was prepared in 50 mM MOPS, 20% glycerol, 500 mM NaCl and free histidine. 10 μL aliquots of 4.8 mg/mL stock solutions of NAR I was stored at –20 °C prior to use. The solution of NAR I was denatured 1:5 with SDS sample buffer containing 2-mercaptoethanol and heated to 70 °C for 5 minutes before loading onto the gel. Gel electrophoresis was performed with 4%/12% SDS-PAGE (1mm thick) mini-gels and Coomassie blue stained. The excised gel piece was digested with trypsin and CNBr/trypsin and analysed by MALDI MS as described below.

2.2.4 ER Membrane Proteins

A complex mixture of proteins extracted from the ER of the liver from healthy mice was prepared by Dr. Barbara Knoblach of Professor Marek Michalak's research group at the Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada.[16] The minced liver was homogenized at 4,000 rpm in 0.5 M sucrose, 1% dextran, 5 mM $MgCl_2$, 37.5 mM Tris-maleate at pH 6.5. The mixture was centrifuged at 5,000 g for 15 minutes and the supernatant collected. The supernatant was then diluted 1:4 with 0.5 M sucrose buffer and spun down at 8,500 g for 5 min. The resulting supernatant was collected and subjected to sucrose density gradient fractionation. Sucrose gradients of 1.3 M, 1.5 M and 2.0 M were added to the supernatant and spun at 90,000 g for 90 minutes. The ER was isolated and washed twice with 10 mM Tris-HCl at pH 7.4. This complex mixture of proteins was first extracted with Triton X-114 to isolate the hydrophobic proteins (fraction T). The remaining extract was solubilized in urea buffer and separated on a 2D gel. The remaining urea-insoluble pellet was further dissolved with SDS (fraction U). Fractions T and U were both separated on a 1D gel and stained by Coomassie Blue. The resultant 1D gel should contain the membrane proteins or hydrophobic proteins of the ER and it is to these bands that we applied the CNBr/trypsin digestion method for purposes of hydrophobic protein identification.

The excised gel pieces from the ER sample were digested with trypsin and CNBr/trypsin and analysed by MALDI MS as described below. LC-ESI MS was also performed. The LC separation was performed on a Waters CapLC, using a water/acetonitrile (0.2% formic acid) gradient, on a PicoFrit capillary column (New Objectives, Woburn, MA) (BioBasic C18, 5 micron particle size, 10 cm x 75 μm ID, 15 μm tip). The eluted peptides were then electrosprayed and analyzed on a MicroMass Q-TOF 2 using automated data dependent MS to MS/MS switching. The resultant MS/MS data was then searched against MASCOT (Matrix Sciences) search engine to perform the protein database search and identification.

2.2.5 In-gel Digestion

The proteins of interest were excised and subjected to an in-gel CNBr/trypsin digestion. Reduction and alkylation steps were performed prior to CNBr digestion for proteins containing cysteine residues. Reduction was performed at 56 °C for at least 30 min in 10 mM DTT in 100 mM NH_4HCO_3. Alkylation was performed in the dark at room temperature for 30 min in 55 mM iodoacetamide in 100 mM NH_4HCO_3. CNBr digestion was performed overnight (~12 hr) with 100 mM CNBr in 50% TFA in the dark at room temperature.[17] The CNBr digestion was quenched the next day by washing the gel piece with 1 mL of water and evaporating off the CNBr and TFA by speed vacufuge. Note that the washing solution was not discarded; it was dried along with the gel piece. In one study, the spectra of BR digest with and without discarding of the wash solution were obtained and compared. It was found that the resulting mass spectra were identical. However, drying the gel piece containing the wash solution did not affect the subsequent experiments. Therefore, to avoid potential peptide loss during washing, we did not discard the wash solution. The washing step is necessary to rid the gel of excess CNBr and to evaporate the TFA from the gel piece. The wash ensures that the protein/peptides will be in a basic environment for trypsin digestion. To speed up this process, concentrated ammonium bicarbonate was used to neutralize TFA and the resulting byproducts of carbon dioxide and ammonia gases simply evaporate off. A typical in-gel trypsin digestion protocol[18] was then applied. A solution of 12.5 ng/μL trypsin in 50mM NH_4HCO_3 was added just enough to cover the gel piece. The gel piece was then smashed into small pieces to increase the surface area of proteins/peptides in the solution for improved trypsin accessibility. CNBr is a small chemical reagent and can readily diffuse into the gel piece. Smashing the gel piece at the CNBr digestion stage would make it difficult in subsequent sample handling steps. Trypsin is a large molecule and in our case, smashing the gel piece seems to give better mass spectrometric results in terms of detection sensitivity, sequence coverage, and reproducibility. In the literature there are many reports that do not smash the gel piece for trypsin digestion and the sensitivity and peptide coverage seem to be comparable to those obtained with gel smashing. There is really no "theory" to explain these observations.

To enhance peptide solubility and digestion efficiency, a 0.005% of *n*-octyl glucoside was added to the trypsin digestion buffer. Trypsin digestion was performed for 3 hr at 37 °C. Sequential extraction of peptides from the gel was done twice with 0.25% TFA/20% ACN, twice with 0.25% TFA/50% ACN and once with 0.25% TFA/80% ACN. The extracts were pooled together and the organic solvent dried off by speed vacufuge.

2.2.6 Sample Preparation for MALDI MS Analysis

The dried peptide extract was diluted with 5 μL 0.1% TFA and desalted with μ-C18 ZipTips. The peptides bound to the ZipTip were eluted out sequentially in 5 μL each of 20%, 50% and 80% ACN in 0.1% HAc (acetic acid) solution or just 50% ACN/0.1%HAc. The sequentially eluted peptide solution was dried to reduce its volume to about 5 μL and a matrix solution was added (see below). The purpose of using the ZipTip is to remove salts from the samples prior to MS detection, but it does not completely remove the n-OG, a non-ionic detergent. The two-layer sample deposition method[19] with 4-HCCA as matrix was used in the MALDI MS analysis. The first layer was prepared as a 20 mg/mL 4-HCCA solution in 20% methanol/acetone HCCA and the second layer with a saturating solution of matrix in 30% (v/v) methanol/water. The second layer was added to the Zip-Tipped peptide mixture to a ratio of matrix to analyte of 4:1 and the mixture vortexed. After 0.5 μL of the first layer was deposited on the sample probe and air-dried, 0.5 μL of the second layer was deposited on top of the first layer, allowed to air dry and washed twice with 1 μL water.

2.2.7 MALDI Mass Spectrometry

MALDI MS was performed on a Bruker Reflex III Matrix-Assisted Laser Desorption Ionization time-of-flight mass spectrometer (Bremen/Leipzig, Germany) equipped with a SCOUT 384 multiprobe inlet and a 337 nm nitrogen laser operated with a 3 ns pulse in positive ion mode with delayed extraction using reflectron mode. The sample spot was scanned with the laser beam under video observation and spectra were acquired by averaging 300-500 individual laser shots and processed with the Bruker supporting software. The spectra were internally calibrated with trypsin autolysis peptide peaks and matrix peaks. The data were then reprocessed using the Igor Pro software package (WaveMetrics, Lake Oswego, Oregon, USA). Each spectrum was normalized to the most intense signal in the mass range displayed.

2.2.8 MALDI Tandem Mass Spectrometry (MS/MS)

Selected peptides were fragmented on an Applied Biosystems/MDS-Sciex QSTAR Pulsar QqTOF instrument equipped with an orthogonal MALDI source employing a 337 nm nitrogen laser (Concord, ON, Canada). The instrument was operated in positive ion mode and collision-induced dissociation (CID) of peptides was achieved with argon as collision gas. Spectra were acquired and processed using Sciex supporting software and re-processed with Igor Pro software.

2.2.9 Mass Spectra Interpretation and Database Searching

Known contaminant peak masses were eliminated for each spectrum. The sample peaks were determined by comparing the peaks from a sample to the peaks resulting from a blank piece of gel. Only sample peaks were considered for database searching. The artifactual modifications of peptides by electrophoresis, such as acrylamide adducts to cysteine and oxidation of methionine, were also considered for the database searching. Both peptide mapping and the peptide sequencing results were searched for protein identification using the Mascot search program (http://www.matrixscience.com) and the UCSF Protein Prospector Database (http://prospector.ucsf.edu). The obtained partial sequence information for each peptide was used to confirm, discard, or correct the previously obtained results from the peptide map search.

2.3 Results and Discussion

2.3.1 Development of CNBr/Trypsin Digestion Method using Bacteriorhodopsin

Trypsin is a common choice of enzyme used in the digestion of hydrophilic proteins. It cleaves at the C-terminus of lysine (K) and arginine (R) residues except when the next residue is proline. However, these hydrophilic amino acid residues, K and R, are rarely found in the transmembrane region of membrane proteins. Cyanogen bromide (CNBr) cleavage at methionine (M) residues, which is more abundant in the transmembrane regions than R or K residues, is often required for samples that are water-insoluble or for membrane proteins where enzymes have poor accessibility for proteolytic attack.[20] CNBr electrophilically reacts with sulfur from methionine residues except when the methionine residue is oxidized. Following the hydrolysis reaction, a peptide is cleaved with a new amino terminus and homoserine and/or homoserine lactone (h) results in the C-terminal position of the remaining peptide. Due to the low number of M found in proteins, CNBr cleavage produces a small number of large peptide fragments with MW typically > 2000 which are difficult to extract from gel pieces. To produce a larger number of smaller peptides than that obtained by using CNBr alone, trypsin can be used to further digest the sample.

In this work, bacteriorhodopsin (BR) was used as a model to develop and optimize the CNBr/trypsin digestion protocol for the analysis of hydrophobic membrane proteins. BR is a very hydrophobic protein with seven transmembrane (TM) domains (see Figure 2.1). BR contains seven R, seven K and nine M residues. A typical BR in-gel tryptic digestion generated only one possible BR specific peptide peak in the MALDI mass spectrum based on peptide mass matching, as indicated in Figure 2.2A. Other peaks are from trypsin and keratin peptides as well as background peaks from the gel or matrix clusters. MALDI MS/MS of the peptide ion at m/z 2162.0 did not yield any fragment ions, likely due to the low intensity as well as the high mass of this peptide ion. Even if this m/z 2162.0 peak were from BR with a sequence of (R)SRAIFGEAEAPEPSAGDGAAATS(-), it would only give a sequence coverage of 9%. The solid underlined sequence in Figure 2.1 indicates the location of this probable trypsin-cleaved peptide in the amino acid sequence of BR. This pictorial representation of the sequence coverage shows the recovered peptide is not found in the transmembrane regions of BR. This may be attributed to the low digestion efficiency of trypsin in these regions to generate a sufficient amount of peptides to be detected and poor extraction of large hydrophobic peptides generated from trypsin digestion. Evidence of the latter has been reported by van Montfort *et al.* for another membrane protein.[15]

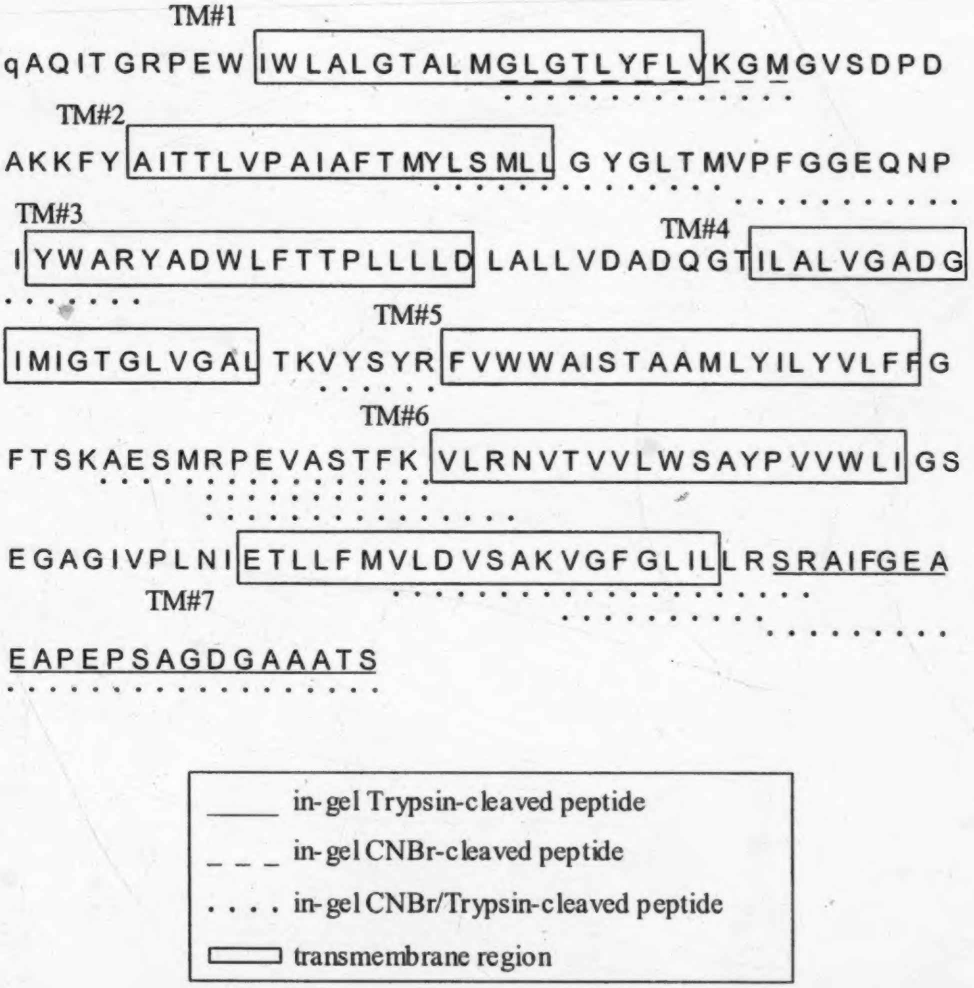

TM#1
qAQITGRPEW IWLALGTALMGLGTLYFLV KGMGVSDPD
TM#2
AKKFY AITTLVPAIAFTMYLSMLL GYGLTMVPFGGEQNP
TM#3
I YWARYADWLFTTPLLLLD LALLVDADQGT
TM#4
ILALVGADG
IMIGTGLVGAL TKVYSYR
TM#5
FVWWAISTAAMLYILYVLFF G
FTSKAESMRPEVASTFK
TM#6
VLRNVTVVLWSAYPVVWLI GS
EGAGIVPLNI ETLLFMVLDVSAKVGFGLIL LRSRAIFGEA
TM#7
EAPEPSAGDGAAATS
in-gel Trypsin-cleaved peptide
in-gel CNBr-cleaved peptide
in-gel CNBr/Trypsin-cleaved peptide
transmembrane region

Figure 2.1 Bacteriorhodopsin sequence showing the seven transmembrane regions (indicated by a retangular box). The solid underlined sequence of BR shows the cleaved peptide of an in-gel trypsin digestion of 1 μg BR. The dashed underlined sequence of BR shows the cleaved peptide of an in-gel CNBr digestion of 5 μg BR. The dotted underlined sequence of BR shows the cleaved peptides of an in-gel CNBr/trypsin digestion of 5 μg BR.

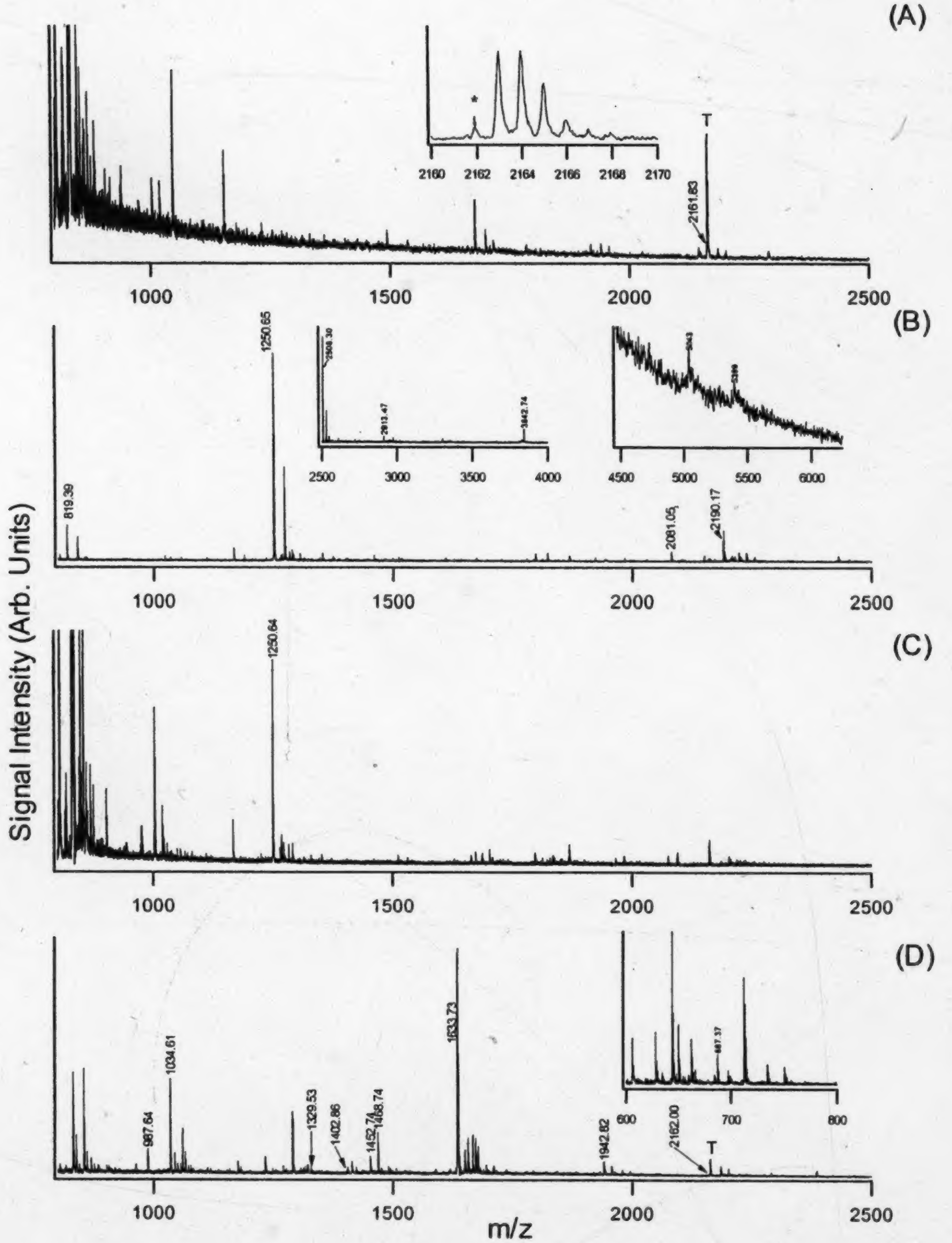


Figure 2.2 MALDI mass spectra of BR using various digestion protocols: (A) In-gel trypsin digest of 1 μg BR, (B) In-solution CNBr digest of 20 μL of 1 μg/μL BR, (C) In-gel CNBr digest of 5 μg BR and (D) In-gel CNBr/trypsin digest of 5 μg BR. The peaks labeled with the corresponding m/z value are from the peptides of

bacteriorhodopsin. The masses of the peaks shown in the inset in (B) were determined by external calibration. The peak labeled with 'T' is the trypsin peptide (LGEDNINVVEGNEQFISASK) at m/z 2163.05.

To obtain peptides in the transmembrane regions not accessible to trypsin, CNBr was used to selectively cleave at M residues. The in-solution CNBr digestion of 20 μL of 1 μg/μL BR (see Figure 2.2B) resulted in detection of specific CNBr-digested peptides (labeled with the corresponding m/z value) with a total sequence coverage of 100%. By comparison, in-solution trypsin digestion of the same amount of BR resulted in only one confirmed peptide at m/z 1452.4 which corresponds to BR peptide 160-172 (spectrum not shown). No other BR peptides were found. It should be noted that the extent of sequence coverage generated by a given digestion method is dependent on the amount of the sample used. For a low efficiency digestion, such as trypsin digestion of BR, increasing the sample amount can generate more peptides that have sufficiently high concentrations to be detected by MS, resulting in better sequence coverage.[21] Compared to trypsin digestion of the same amount, the complete recovery of BR peptides by CNBr in-solution demonstrates the advantage of CNBr to cleave peptides within the transmembrane regions for successful peptide recovery and detection by MS. It should be noted that, in our experiments of dealing with picomoles of protein, it is not clear whether CNBr cleavage is complete or not. Failure to detect the expected CNBr-cleaved peptides does not mean that the CNBr cleavage is not complete. When the working amount of sample is small, not every peptide generated from the digestion process can be detected, many peptides may be produced in such a small amount that MS just cannot detect them. Next, the performance of CNBr was tested using the in-gel approach.

In-gel CNBr digestion of 5 μg BR followed by MALDI MS peptide mapping was not as successful (see Figure 2.2C). This is likely due to the reduction of in-gel digestion efficiency compared to in-solution digestion, and/or difficulty of extracting digested peptides from the gel at this level of sample loading. Shown in Figure 2.2C, only one CNBr-cleaved peptide is recovered. The dashed underlined BR peptide sequence (M)GLGTLYFLVKGh(G) at m/z 1250.7 shown in Figure 2.1 represents a sequence coverage of 5%.

The MALDI mass spectrum obtained from sequential CNBr/trypsin digestion of 5 μg BR is shown in Figure 2.2D. Figure 2.2D clearly shows that CNBr/trypsin digestion is advantageous over single-cleaving reagent digestion methods for in-gel. The peptide mass map obtained has a sequence coverage of 35%. The coverage is a result of the detection of specific CNBr-trypsin digested peptides with high intensity of some peptides while other peptides expected from the known cleavage sites are not detected, likely due to many contributing factors including decreased digestion efficiency on these sites, reduced peptide recovery from the gel, sample loss during sample workup, and ion suppression in MS detection. It is worthy to note that the peptides recovered from the CNBr/trypsin digestion are indeed from a tryptic cleavage of CNBr-fragmented peptides. In Figure 2.1, the peptides (dotted underlined sequences) arise from almost all seven transmembrane regions of BR except for TM #4 and #5, and were cleaved at an arginine or lysine residue and/or at a methionine residue. This illustrates the usefulness of CNBr to cleave the hydrophobic protein into smaller peptides, allowing trypsin to cleave at the R and K residues of these CNBr-derived peptides and subsequently allowing these peptides to be effectively recovered and detected by MS.

2.3.2 Effect of *n*-Octyl Glucoside on MALDI MS Detection

Throughout the entire sample handling process, the aggregation of proteins or of the resulting hydrophobic peptides must be minimized. It has been reported that the addition of a detergent in the digestion buffer can aid in the recovery of extracted peptides.[7,22] Katayama et al. have shown that the addition of a non-ionic detergent, *n*-octyl glucoside (*n*-OG), enhances the digestion efficiency by increasing the solubility of proteins/peptides and by preventing the adsorption of peptides on the walls of sample vials and pipette tips.[7] van Montfort *et al.* speculated that *n*-OG improved peptide solubility during the resolubilization and crystallization step.[22] In general, the addition of *n*-OG to the digestion mixture resulted in the preferential detection of the MALDI-MS response of larger peptides, whereas it lead to the suppression of the lower masses.[23] In the development of an optimized procedure, *n*-OG was used in the digestion mixture to enhance the digestion efficiency and to give optimized peptide recovery of BR peptides.

Since the interest of our work lies in detecting sub-microgram amounts of starting hydrophobic protein and the critical micellar concentration of *n*-OG is 23 mM, the proper amount of *n*-OG must be carefully determined to reduce suppression of the peptide mass signals. The addition of 0.02% *n*-OG to the digestion buffer of 500 ng BR resulted in the preferential detection of the higher mass peptides (above m/z 1000), while suppression was observed for the peptides with lower masses (e.g., BR peptides with m/z 687.3 and 987.6) (data not shown). Overall, under the conditions employed, addition of 0.005% *n*-OG was shown to optimize peptide recovery while exhibiting minimal suppression of lower mass peptides. We did not find any improvement in signal strength in using higher amounts of n-OG up to 0.1%, but the spectra obtained were less reproducible from one experiment to another and from one operator to another, compared to the use of 0.005%. Also tested was the effect of *n*-OG in the digestion buffer, in the extraction buffer and/or in the MALDI MS sample spot. However, the results are not conclusive since peptide solubility depends on many factors. The role of *n*-OG remains unclear and a more detailed look is beyond the scope of this study.

2.3.3 Detection Sensitivity of CNBr/Tryptic Digest of BR In-Gel

The usefulness of the CNBr/trypsin in-gel approach is illustrated in its ability to detect low abundant and therefore, a low amount of starting proteins. Figure 2.3 illustrates the detection sensitivity of this in-gel digestion method. It is sensitive enough to analyze 200 ng BR with a sequence coverage of 20%, as shown in Figure 2.3A. With a sample loading of 500 ng and addition of 0.005% *n*-OG to the digestion buffer, several peptides are detected with high intensity (see Figure 2.3B), which allows the acquisition of good quality MS/MS spectra for protein identification. A representative MALDI MS/MS spectrum of a peptide digested by CNBr/trypsin, the peptide at m/z 1034.5 with sequence (M)RPEVASTFK(V), is shown in Figure 2.3C. The five peaks at m/z 687.3, 987.6, 1034.5, 1468.6 and 1633.7 shown in Figure 2.3B were confirmed by MS/MS to be from BR. The peak at m/z 1110.5 has its mass matching with one of the expected BR peptides. Some common peaks from trypsin and keratin peptides are labeled and others are likely from background peaks of unknown origins. For comparison, a representative spectrum obtained from trypsin digestion of 500 ng BR followed by CNBr, according to the protocol of reference 15, is shown in Figure 2.3D. The overall spectral quality in terms of the signal-to-noise ratio and the number of BR peaks detected for Figure 2.3B is superior. In the case of trypsin/CNBr digestion, MS/MS spectrum can only be obtained from the peak at m/z 987.6. Moreover, it is our experience that the CNBr/trypsin protocol reported in this work is more reproducible in repeat experiments for different operators. Our rationale in choosing this order of digestion was that the chemical cleaving reagent such as CNBr is expected to be more readily interacting with a membrane protein in gel, compared to a protease which is bulky and must retain its optimal conformer for activity. After CNBr digestion, the smaller fragments are expected to be more accessible to trypsin for further digestion. Our data appears to support this

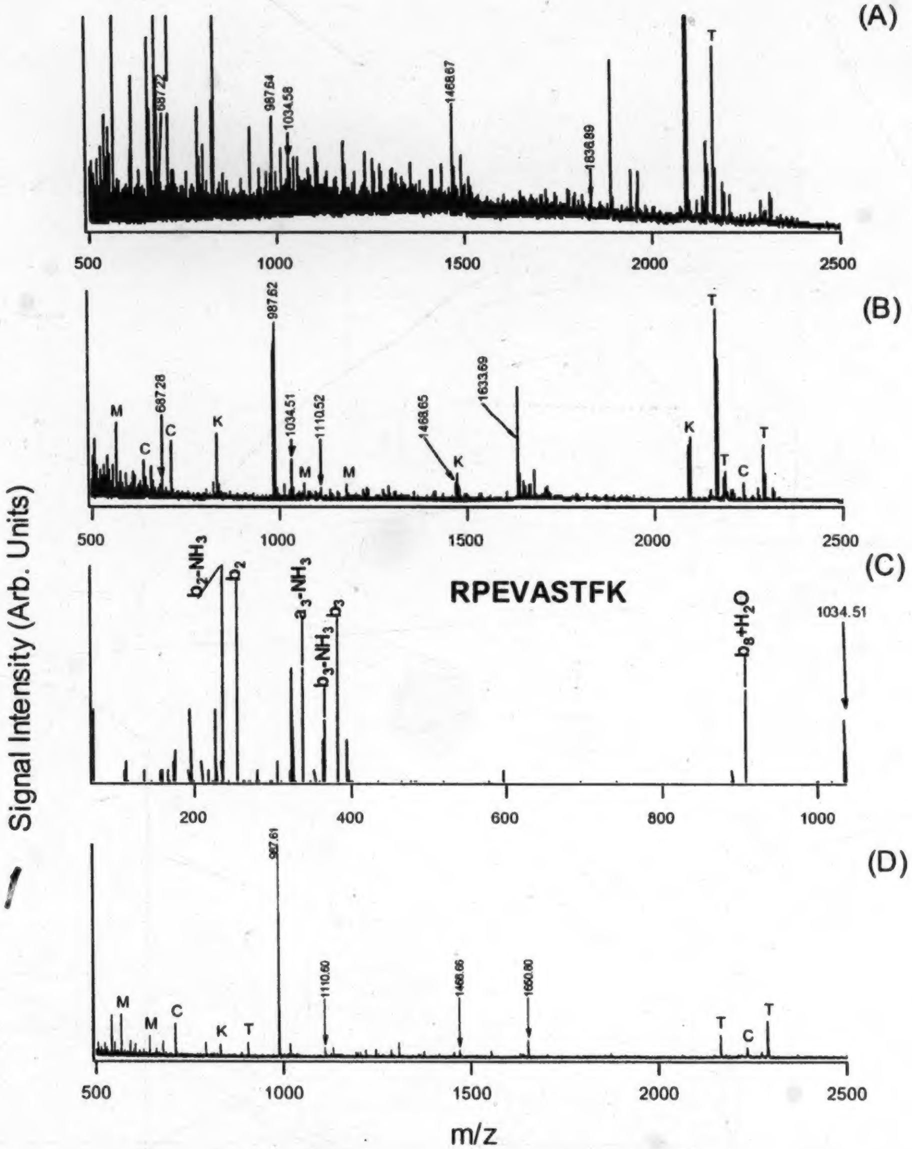


Figure 2.3 MALDI mass spectra of in-gel CNBr/trypsin digestion of BR: (A) 200 ng BR in-gel and (B) 500 ng BR in-gel using 0.005% n-OG in digestion buffer. (C) MALDI QqTOF MS/MS spectrum of in-gel CNBr/trypsin digestion of 500 ng BR peptide (RPEVASTFK) at m/z 1034.5. (D) MALDI mass spectrum of an in-gel trypsin/CNBr digestion of 500 ng BR. The peaks labeled with M, C, K, and T originate from the HCCA matrix, common contaminant peaks of in-gel experiments, keratin and tryptic peptide peaks, respectively.

hypothesis.

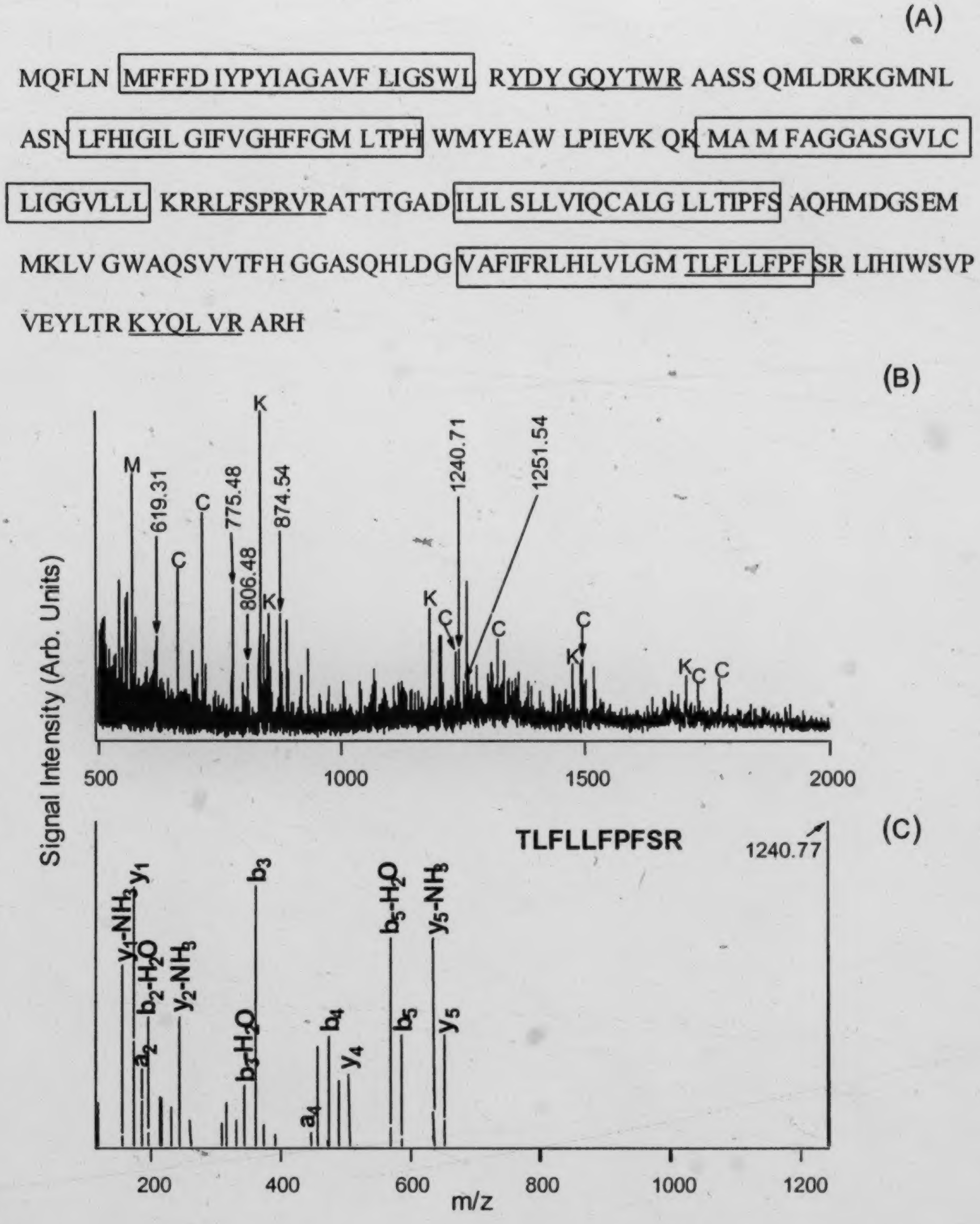
It should be noted that, compared to the analysis of less hydrophobic proteins where trypsin in-gel digestion can be used, the CNBr/trypsin digestion is about 5 to 10 times less sensitive. This sensitivity reduction is likely resulted from the decrease of digestion efficiency of membrane proteins, as well as the decrease of extraction efficiency, sample recovery and ionization efficiency of the resulting peptides. Nevertheless, sub-microgram detection is adequate for many real world proteomics applications and is comparable to the detection limit of using Coomassie staining of membrane proteins in which a distinct band is visible with the naked eye.

2.3.4 Source of Unrecovered Peptides from a MALDI MS Analysis of an In-Gel CNBr/Tryptic Digest

To compare the results shown in Figure 2.2, several reasons may account for unrecovered CNBr/tryptic peptides since sample loss is typical for in-gel experiments. Some peptides obtained in the CNBr-only in-solution digest (Figure 2.2C) were not observed in the sequential CNBr/trypsin in-gel digest because of sample loss during successive steps following CNBr digestion, such as during the washing of the gel piece with water. Also, some CNBr-cleaved peptides may still be fairly hydrophobic and therefore are more difficult to extract from the gel than for the in-solution CNBr digest. The use of a high amount of acid in the extraction solvent system such as 70% TFA has been suggested for extracting hydrophobic peptides from a gel.[22] However, for the BR peptides, we did not see any additional peaks from the gel with the 70% TFA extraction after it has been extracted by using our protocol described in the Experimental section. The common modification observed for the BR-specific peptides is oxidation of methionine to the sulfoxide form; however, this was not observed in the peptide mass map obtained. Autodigestion products of trypsin and unassignable peaks in the MALDI mass spectra of peptide maps from gels is common.[18] Some peaks may not correspond to predicted peptide masses or from possible post-translational modifications. Human keratins which accumulate from chemicals and/or sample handling often become ubiquitous at these low levels and may suppress the detection of the less abundant peptides specific to the protein of interest. The emergence of keratin peptides in a mass spectrum is unpredictable and cannot be subtracted from a control.

2.3.5 Application to a Membrane Protein with Cysteine Groups: NAR I E. Coli

To validate the usefulness of our method, Nitrate reductase 1 gamma chain (NAR I *E. coli*), a 25.5 kDa recombinant transmembrane protein, was studied (see Figure 2.4). The gamma chain is a 5-transmembrane-embedded heme-iron unit. NAR I contains cysteine groups and reduction and alkylation was necessary before digestion. However for this particular membrane protein, the CNBr/tryptic peptides (labelled with the corresponding m/z value) obtained were comparable with those from trypsin-only digestion except for two peptides at m/z 1240.7 and 1251.5. All other peptides observed were below m/z 1000, the region where matrix clusters dominate in the spectrum. The peptides at m/z 1240.7 and 1251.5 are specific to NAR I, which were confirmed by MS/MS (data not shown). The peptide at m/z 1240.7 has the sequence (M)TLFLLFPFSR(L) arising from a transmembrane region in which CNBr cleavage at the M residue generated a large peptide with threonine (T) at the N-terminus. Next, tryptic cleavage at the R residue (C-terminus of resultant peptide) cuts the large peptide into a smaller peptide. The smaller peptide can be more easily extracted and recovered for MS detection. The reason for not observing the peptide at m/z 1251.5 (sequence (R)YDYGQYTWR(A)) in the trypsin-only digest, even though the peptide is cleaved by trypsin on both ends, is most likely from the key role of CNBr prior to trypsin cleavage. CNBr can access the hydrophobic sites and cleave at M residues, generating smaller peptides for proteolytic attack by trypsin. This example illustrates that, for membrane proteins that can be digested by trypsin-only, the CNBr/trypsin digestion method can



(A)

MQFLN MFFFD IYPYIAGAVF LIGSWL RYDY GQYTWR AASS QMLDRKGMNL

ASN LFHIGIL GIFVGHFFGM LTPH WMYEAW LPIEVK QK MA M FAGGASGVLC

LIGGVLLL KRRLFSPRVRATTTGAD ILIL SLLVIQCALG LLTIPFS AQHMDGSEM

MKLV GWAQSVVTFH GGASQHLDG VAFIFRLHLVLGM TLFLLFPF SR LIHIWSVP

VEYLTR KYQL VR ARH

Figure 2.4 (A) Sequence of NAR I showing the penta transmembrane regions with CNBr/trypsin cleaved peptides (solid underline). (B) MALDI mass spectrum of in-gel CNBr/trypsin digestion of 1 mg NAR I. (C) MALDI QqTOF MS/MS spectrum of in-gel CNBr/trypsin digestion of 1 mg NAR I peptide (TLFLLFPFSR) at m/z 1240.8. The peaks labeled with the corresponding m/z value is specific to the peptides of NAR I.

provide additional peptides for the purpose of protein identification.

2.3.6 Application to a Complex Biological Extract: ER Membrane Proteins (Mouse)

The application of this method to the characterization of a biological sample is illustrated for the analysis of a complex protein mixture extracted from the endoplasmic recticulum (ER) membrane of mouse liver. As described in the Experimental, the membrane proteins or hydrophobic proteins of the ER were separated by 1D gel. The bands from the 1D gel were excised in two halves; one half of the band was digested with trypsin-only, while the other half of the band was digested with CNBr/trypsin. To illustrate the usefulness of this CNBr/trypsin digestion method for the analysis of complex biological samples as well as the processes involved in data interpretation, we present here the results for Band U7, a ~70 kDa band from the urea-insoluble fraction. The comprehensive proteome map of ER and their biological functional implications will be reported elsewhere.[16]

The MALDI peptide mass maps (see Figure 2.5) generated from both trypsin and CNBr/trypsin digestions were first entered into the database and searched for protein identification. The search results were ambiguous in protein identification of Band U7. The complexity of the peptide mass map is likely due to the presence of a mixture of proteins residing in the band. Therefore, MS/MS of selected peptides was needed to generate sequence information in order for confident identification of the parent proteins.

Figure 2.5 shows that the major differences in the MALDI mass spectra generated from trypsin digest and CNBr/tryptic digest are the relative intensities of the peptide peaks. In particular, the two peptide peaks at m/z 1022.5 and 2080.0 have a significantly higher intensity in the CNBr/trypsin digest compared to the trypsin-only digest. As it is shown in Figure 2.6A, good quality MALDI MS/MS spectrum from the peptide at m/z 1022.5 can be obtained from the CNBr/trypsin digest and this fragment ion spectrum matches a peptide with a sequence of (R)ALHDQLGLR(Q) from very long chain acyl CoA synthetase (ligase), a 70,367 Da peroxisomal integral membrane protein (SwissProt accession# O35488 (Mus musculus)) with three transmembrane domains. However, poor MS/MS spectra were obtained for the peptide at m/z 2080.0 even from the CNBr/trypsin digest. This may be attributed to inefficiency of fragmenting this large peptide ion with the current QqTOF mass spectrometer. A number of other peptide ions detected in both trypsin and CNBr/trypsin digests were selected for MALDI MS/MS. Unfortunately searching these MS/MS spectra against the database did not identify any other proteins with high confidence (i.e., each spectrum produced a number of possible candidates with similar low scores in MASCOT). Usually, the MALDI-QqTOF instrument can reliably identify all of the components in a mixture that contains only a few components whose relative abundances differ by a factor of ~10. More complex mixtures may be investigated by LC/ESI-MS with on-line MS/MS capabilities.[24]

We then proceeded to obtain LC-ESI QTOF MS/MS data with the hope that unique as well as complementary sequence information could be generated from the digests of Band U7. With the ESI MS/MS data, another protein, similar to fatty acid Coenzyme A ligase (long chain 5) (76,206 Da, SwissProt accession# Q8JZR0 (Mus musculus)), was identified from the two peptides at m/z 1052.5 (K)IGFFQGDIR(L) and 2080.0 (K)LAQGEYIAPEKIENVYSR(S). While these two peptides were also detected in the MALDI mass spectra (see Figure 2.5), the quality of their MALDI MS/MS spectra

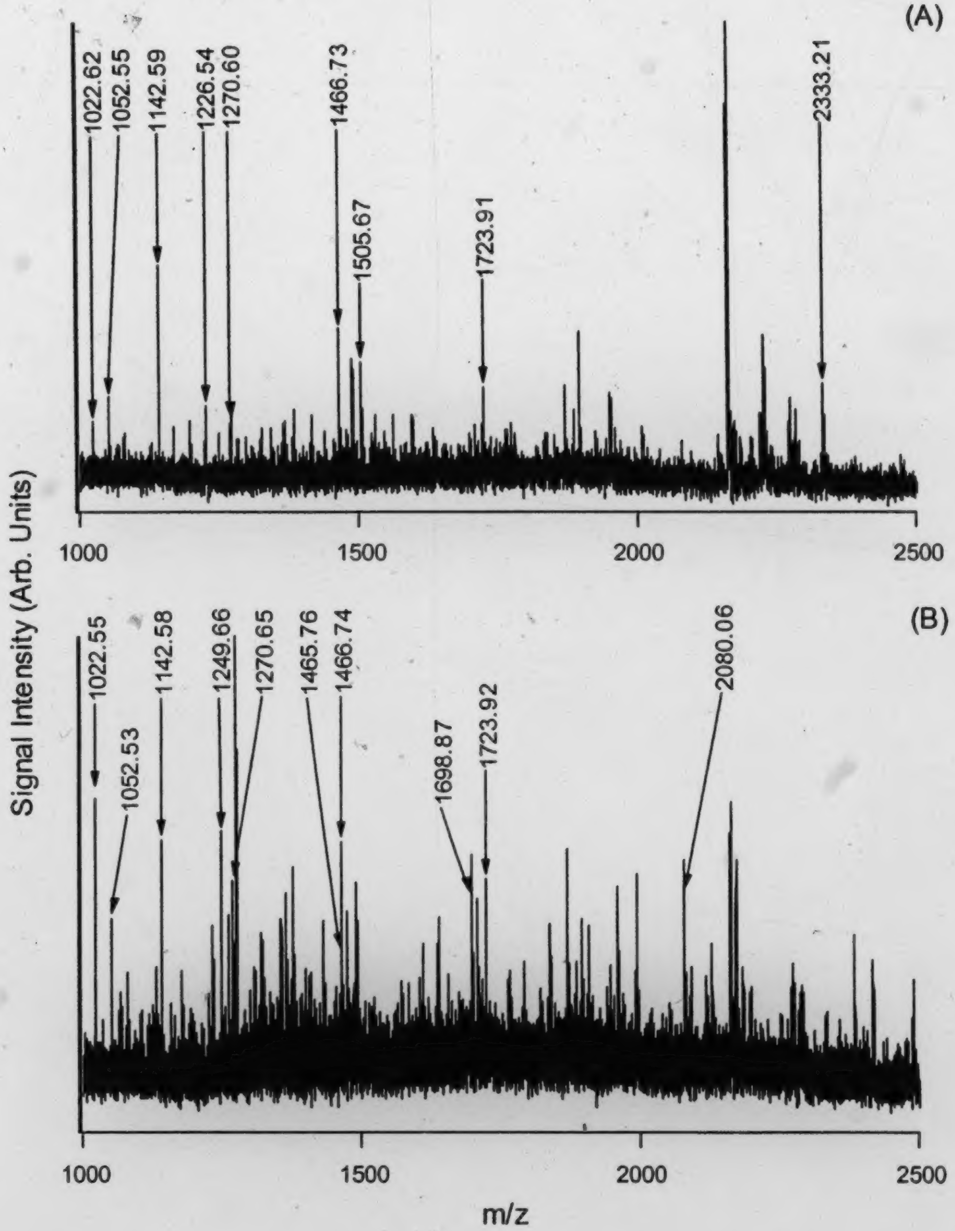


Figure 2.5 MALDI mass spectra of Band U7 from the protein extract of the ER membrane of mice. (A) In-gel trypsin digestion (B) In-gel CNBr/trypsin digestion.

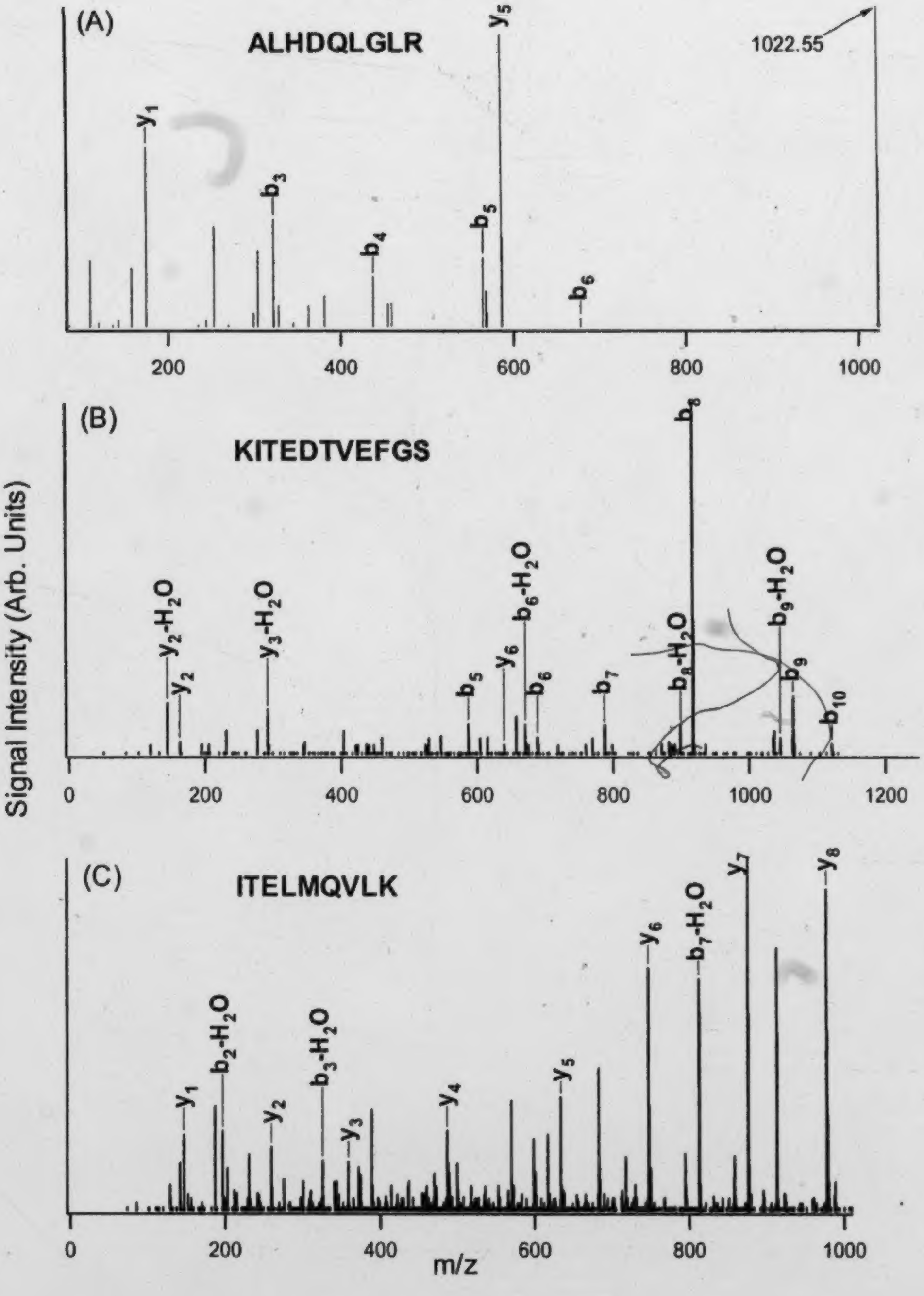


Figure 2.6 MALDI QqTOF MS/MS spectrum of the ligase peptide: (A) peptide (ALHDQLGLR) at m/z 1022.5 in the CNBr/trypsin digest of Band U7. ESI QTOF MS/MS spectrum of the ATP-binding protein: (B) peptide

(KITEDTVEFGS) at m/z 1225.6 in the trypsin digest of Band U7. (C) peptide (ITELMQVLK) at m/z 1090.6 in the CNBr/trypsin digest of Band U7.

was not sufficient for protein identification. However the fragment ions observed in the ESI MS/MS spectra matched with the sequences of these two peptides, thus providing another dimension of confirmation.

Another protein identified in Band U7 is an ATP-binding protein, a 75,482 Da peroxisomal penta transmembrane protein (SwissProt accession# P55096 (Mus musculus)). Tentative identification of this protein was from the ESI MS/MS data where the MS/MS spectrum from a peak at m/z 1225.6 in the tryptic digest correlated well with the peptide sequence, (K)KITEDTVEFGS(-), belonging to the ATP-binding protein (Figure 2.6B). This peptide was not observed in the CNBr/tryptic digestion ESI mass spectrum. To enhance the confidence in protein identification of database search results, it is desirable that the sequence for at least two peptides be characterized. Affirmation that the ATP-binding protein is indeed present in this complex protein mixture was possible only from the peptide at m/z 1090.6 observed in the CNBr/tryptic digest from the ESI MS/MS data (Table 2.1). The peptide at m/z 1090.6 was identified as belonging to the ATP-binding protein with sequence (R)ITELMQVLK(D) + 1 Met-Ox (Figure 2.6C). This peptide at m/z 1090.6 has an oxidized methionine residue, which means that the CNBr was blocked from reacting with the sulfur of the methionine residue. The sample may be oxidized during various steps in the sample preparation and analysis. This oxidation may explain the lack of CNBr-cleaved peptides observed in the digest mixture. However, because this peptide was only observed in the ESI mass spectrum of the CNBr/trypsin digestion, it is clear that the CNBr/trypsin digestion method is useful for providing key peptide information in addition to that obtained from the trypsin-only digestion method.

2.3.7 Identification of ER Membrane Proteins

The peak assignments for proteins identified in Band U7 using digestion by trypsin and CNBr/trypsin and detection by MALDI- and ESI-MS and MS/MS are summarized in Table 2.1. In this case, MS/MS spectra were collected for as many peptide ions as possible. Accurate mass measurement of the peptide mass can also provide useful information for protein confirmation. As shown in Table 2.1, matched peptides have an error mostly below 30 ppm, which is acceptable for the instrumentation employed. Although the sequence coverage increased for both ligase and long chain 5 proteins, no CNBr-cleaved peptides were detected using ESI- and MALDI-MS. However, CNBr-cleaved peptides are observed for the ATP-binding protein. Two CNBr-cleaved peptides from the ATP-binding protein were observed by ESI-MS at m/z 1144.6 (h)LLRMSQALGR(I) and m/z 1381.6 (h)LVSRTYCDVWh(I). Of particular interest is the peptide at m/z 1381.6 (93-103 AA), which is within the transmembrane region (TM1: 84-104 AA). The peptide mass map obtained from the ESI MS data of the CNBr/trypsin digest had a significantly higher sequence coverage of the ATP-binding protein compared to the trypsin-only digest. The total sequence coverage is 8% for the CNBr/trypsin digest and 3% for the trypsin digest, with a combined sequence coverage of 11%.

The purity of the ER sample was verified[16] and all three proteins identified in Band U7 are peroxisomal membrane proteins. Peroxisomes function to rid the body of toxic substances and are relatively more abundant in the liver where toxic byproducts of metabolic reactions accumulate. Thus, the presence of these peroxisomal membrane proteins in Band U7 is consistent with their functions since the proteins were extracted from the ER membrane of the liver. There still remain some peptide signals that could not be assigned. There may be additional proteins whose identification is not possible with this experiment. ESI or MALDI can be used for protein analysis applications, with their own benefits and drawbacks. Advantages of MALDI include higher tolerance to contamination and simpler sample handling. Alternatively, the complexity of mixtures may be reduced by additional off-line separation by other methods, including chromatography and electrophoresis, either before or after digestions and the simplified fractions can be collected onto MALDI target for further analysis. Nevertheless, the use of CNBr/trypsin digestion has allowed us to identify and confirm more proteins than otherwise possible. For example, it is possible to identify the ATP-binding protein with greater confidence with the aid of CNBr/trypsin cleavage.

Table 2.1 Identification and Detection of Peptides from Band U7.

Identity	Theoretical mass (m/z)/ Peptide sequence	Exp.t'l mass (m/z)	Mass accuracy (ppm)	Digestion by Trypsin	Digestion by CNBr/ Trypsin	Detection mode
Very long chain Acyl CoA synthetase (Ligase)	1022.57 (R)ALHDQLGLR(Q)	1022.5	24		√	MALDI MS, MS/MS ESI MS, MS/MS
	1142.58 (K)DAVSVFYVSR(T)	1142.6	4	√	√	MALDI MS, MS/MS
	1226.64 (R)YFLQLANMAR(R)	1226.5	78	√		MALDI MS, MS/MS
	1249.61 (K)YDVEKDEPVR(D)	1249.7	43		√	MALDI MS, MS/MS
	1270.68 (K)KDAVSVFYVSR(T)	1270.6	23	√	√	MALDI MS, MS/MS
	1310.62 (R)DETLTYAQVDR(R)	1310.6	10	√		ESI MS
	1363.57 (K)DTLYFMDDAEK(T)	1363.6	2	√		ESI MS
	1365.73 (R)IQDTIEITGTFK(H)	1365.7	5	√	√	ESI MS
	1465.80 (K)ITQLTPFIGYAGGK(T)	1465.7	31	√	√	MALDI MS, MS/MS ESI MS, MS/MS
	1466.72 (R)DETLTYAQVDRR(S)	1466.7	11	√	√	MALDI MS, MS/MS ESI MS, MS/MS
	1505.72 (R)YLCNTPQKPNDR(D)	1505.7	31	√		MALDI MS
	1529.72 (K)VDGVSAEPTPESWR(S)	1529.7	35	√		ESI MS, MS/MS

	1884.94 (K)TFVPMTENIYNAIIDK(T) + Met-ox	1884.9	22	√		ESI MS
	2333.19 (R)SEVTFSTPALYIYTSGTTGLPK(A)	2333.2	8	√		MALDI MS ESI MS, MS/MS
Similar to fatty acid Coenzyme A ligase (long chain 5)	911.50 (K)FFQTQIK(S)	911.5	10		√	ESI MS, MS/MS
	1014.61 (K)IQGSLGGKVR(L)	1014.6	36	√		ESI MS
	1052.55 (K)IGFFQGDIR(L)	1052.5	22	√	√	MALDI MS, MS/MS (CT only) ESI MS, MS/MS
	1058.57 (R)EAILEDLQK(I)	1058.6	13		√	ESI MS, MS/MS
	1070.53 (K)TLYENFQR(G)	1070.5	35	√		ESI MS, MS/MS
	1131.71 (K)FLLNLAIISK(F)	1131.7	12	√		ESI MS
	1301.66 (K)DPEKTQEVLDK(D)	1301.6	22		√	ESI MS
	1723.96 (R)SRPVLQVFVHGESLR(S)	1723.9	24	√	√	MALDI MS
	1951.07 (R)LMITGAAPISTPVLTFFR (A) + 1 Met-ox	1951.0	27	√		ESI MS
	2059.11 (R)SFLIGVVVPDPDSLPSFAAK(I)	2059.1	1	√		ESI MS
	2080.07 (K)LAQGEYIAPEKIENVYSR(S)	2080.1	6		√	MALDI MS ESI MS, MS/MS

ATP-Binding Protein	1090.62 (R)ITELMQVLK(D) +1Met-ox	1090.6	17		√	ESI MS, MS/MS
	1144.66 (h)LLRMSQALGR(I)	1144.6	20		√	ESI MS
	1169.63 (-)AAFSKYLTAR(N) N-terminal acetylated	1169.7	75	√		ESI MS
	1225.59 (K)KITEDTVEFGS(-)	1225.6	12	√		ESI MS, MS/MS
	1302.74 (K)VGITLFTVSHTK(C)	1302.7	40		√	ESI MS
	1381.66 (h)LVSRTYCDVWh(I)	1381.6	20		√	ESI MS
	1418.81 (K)ERAVVDKVFLSR(L)	1418.8	6		√	ESI MS
	1698.88 (K)EYLDNVQLGHILER(E)	1698.9	7		√	MALDI MS

2.4 Conclusions

We have demonstrated that in-gel CNBr cleavage followed by trypsin digestion is useful for peptide mapping of hydrophobic proteins. The use of *n*-octyl-β-D-glucopyranoside in the digestion buffer can aid in the solubilization of hydrophobic peptides and improve the peptide signals. Bacteriorhodopsin, a model hydrophobic membrane protein, can be identified from as low as 8 pmol starting sample. It was demonstrated that the sequence coverage obtained for BR, as well as the recovery of peptides within the transmembrane region was significantly improved using the CNBr/trypsin digestion method. Application of this sequential digestion method to a complex protein mixture extracted from a biological sample, the endoplasmic reticulum, verified its vital role in detection and subsequent identification of hydrophobic proteins that were not easily characterized from trypsin-only digested peptides.

2.5 Literature Cited

1) Ablonczy, Z.; Kono, M.; Crouch, R.; Knapp, D. *Anal. Chem.* **2001**, 73, 4774-4779.
2) Wu, C.C.; Yates J.R. 3rd. *Nat. Biotechnol.* **2003**, 21, 262-267.
3) Washburn, M.P.; Wolters, D.; Yates J.R. 3rd. *Nat. Biotechnol.* **2001**, 19, 242-247.
4) Blonder, J.; Goshe, M.B.; Moore R.J.; Pasa-Tolic L.; Masselon C.D.; Lipton M.S.; Smith R.D. *J. Proteome Res.* **2002**, 1, 351-360.
5) Han D.K.; Eng J.; Zhou H.; Aebersold R. *Nat. Biotechnol.* **2001**, 19, 946-951.
6) von Haller P.D.; Donohoe S.; Goodlett D.R.; Aebersold R.; Watts J.D. *Proteomics* **2001**, 1,1010-1021.
7) Katayama, H.; Nagasu, T.; Oda, Y. *Rapid Commun. Mass Spectrom.* **2001**, 15, 1416-1421 and references 2 and 3 therein.
8) Ferro, M.; Seigneurin-Berny, D.; Rolland, N.; Chapel, A.; Salvi, D.; Garin, J.; Joyard, J. *Electrophoresis* **2000**, 21, 3517-3526.
9) Santoni, V.; Molly, M.; Rabilloud, T. *Electrophoresis* **2000**, 21, 1054-1070.
10) Luecke, H.; Richter, H-T.; Lanyi, J. K. *Science* **1998**, 280, 1934-1937.
11) Ball, L. E.; Oatis, J. E., Jr.; Dharmasiri, K.; Busman, M.; Wang, J.; Cowden, L. B.; Galijatovic, A.; Chen, N.; Crouch, R. K.; Knapp, D. R. *Protein. Sci.* **1998**, 7, 758-764.
12) Stone, K. L.; McNutty, D. E.; Lopresti, M. L.; Crawford, J. M.; DeAngelis, R.; Williams, K. R. In *Techniques in Protein Chemistry III*; (Angeletti, R. H.,

Ed.); Academic Press: San Diego, 1992; pp 23-24.

13) Sun, Y.; Bauer, M. D.; Lu, W. *J. Mass Spectrom.* **1998**, 33, 1009-1016.

14) Rao, S.; Aberg, F.; Nieves, E.; Horwitz, S. B.; Orr, G.A. *Biochemistry* **2001**, 40, 2096-2103.

15) van Montfort B.A.; Doeven M.K.; Canas B.; Veenhoff L.M.; Poolman B.; Robillard G.T. *Biochim. Biophys. Acta* **2002**, 1555, 111-115.

16) Knoblach, B.; Keller, B. O.; Groenendyk, J.; Aldred, S.; Zheng, J., Lemire, B. D.; Li, L.; Michalak, M. *Molecular and Cellular Proteomics 2003,* in press.

17) Loo, R. R. O.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. *Anal. Chem.* **1996**, 68, 1910-1917.

18) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, 68, 850-858.Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem.* **1996**, 68, 2721-2725.

20) Kellner, R.; Houthaeve, T. In *Microcharacterization of Proteins*; (Kellner, R.; Lottspeich, F.; Meyer, H. E., Eds.); Wiley-VCH: Weinheim, 1999; pp 108-110.

21) Hixson K.K.; Rodriguez N.; Camp D.G. 2nd; Strittmatter E.F.; Lipton M.S.; Smith R.D. *Electrophoresis* **2002**, 23, 3224-3232.

22) van Montfort, B. A.; Canas, B.; Duurkens, R.; Godovac-Zimmermann, J.; Robillard, G. T. *J. Mass Spectrom.* **2002**, 37, 322-330.

23) Cohen, S.L.; Chait, B.T. *Anal. Chem.* **1996**, 68, 31-37.

24) Chernushevich, I. V.; Loboda, A. V.; Thomson, B. A. *J. Mass Spectrom.* **2001**, *36*, 849-865.

Chapter 3

Mass Spectrometric Approach to Protein Profiling of Squamous Cell Carcinoma of the Human Tongue

3.1 Introduction

3.1.1 Cancer Proteomics

Proteomics is the study of the complete protein complement, or the proteome, expressed by a particular cell, tissue, or organism at a set time and environmental conditions.[1] The dynamic nature of the proteome is a consequence of many factors, including differential splicing of the respective mRNAs, post-translational modifications, and temporal and functional regulation of gene expression.[2] At the protein level, the healthy cell is transformed into a neoplastic cell via altered expression, differential protein modification, changes in specific activity, and irregular localization, all of which may affect cellular function. Investigating and identifying these distinct changes is the goal in cancer proteomics.[3] Cancer proteomics is classified into two categories, expression proteomics and functional proteomics.[7] Expression proteomics is the study of the proteins that are differentially displayed in a given tissue, body fluid, or serum. Functional proteomics determines the interaction of the protein in the biological system and its role in biological processes.

The development of tools and methods to help in treatment, detection and prevention of cancer has become an important area of proteomics research. Proteomic technology, such as the commonly employed 2D PAGE and newer techniques of multi-dimensional protein identification technology (MUDPIT) and surface-enhanced laser desorption ionization (SELDI) combined with mass spectrometry, allow for identification of the protein changes caused by the disease process. 2D gel electrophoresis followed by protein identification by mass spectrometry has been the conventional method in proteome analysis.[4,5] The direct comparisons of protein expression and identification of the differentially expressed proteins between normal and tumour tissues for biomarker discovery have been reported in various cancers, such as liver, bladder, lung, esophageal, prostate and breast.[6] The limit of this method lies in the low sensitivity for detection of low abundance proteins. Despite the enhanced separation offered with very narrow range pH IPG strips, only a small percentage of the proteome can be visualized. A 2D gel spot may contain multiple components, and the signals generated, through peptide digests, by ions of low abundance are lost in this approach. The detection of post-translationally modified species is difficult because they are often present in extremely low amounts. Furthermore, not all proteins can be identified through peptide mass mapping alone because many proteins remain to be represented full length in sequence databases. Often, small proteins do not generate sufficient numbers of digestion products for identification. In many cases, additional MS sequencing of peptides is necessary to gain knowledge about sequence identity.[3,7] The detection of a different set of proteins through LC coupled with tandem MS can be used to resolve this problem to some extent. The strength of the tandem mass spectrometer lies in the ability to sequence information for a specific peptide in the presence of multiple peptides in the sample. Reversed-phase LC is used to concentrate and separate peptides from extremely complex mixtures before sequencing by MS.[8,9] LC-MS/MS has been demonstrated to enhance separation and identification of analytes at the low femtomole level.[10]

Yates and coworkers introduced multi-dimensional protein identification technology (MUDPIT).[11,12] In this method, the digest mixture is first fractionated by a strong cation exchanger column, followed by a reversed phase column coupled by a nanoelectrospray device to a MS/MS spectrometer. Identification of ~1500 proteins by this method was reported[12], compared with only a few hundred for 2D gel electrophoresis-based methods. However, in contrast to gel-derived proteome, the MUDPIT-derived proteome can not be quantified or give relative abundance information, but only determine their presence. Although lower abundance proteins in the proteome have been detected, the drawback of MUDPIT is that a large amount of starting protein is required and has very low throughput, even lower than the 2D gel electrophoresis-based approach.[6] For a high throughput approach with inherent advantages and disadvantages, surface-enhanced laser desorption ionization (SELDI) TOF MS[13,14], provides rapid identification of cancer biomarkers and proteomic fingerprinting using on-chip protein fractionation.

Another vital component critical to detection, diagnosis, treatment, monitoring, and prognosis are biomarkers. During the course of a disease process, biomarkers are molecules in a biological system that go through changes, these changes are reflected in the physiologic state, and allows a factor comparable across populations.[3] In a state-of-the-art MS laboratory, the current detection limits are in the low subpicomole range required for accurate protein identification. The protein sequence must be present in the database to be identified, although only a small fraction of known human proteins have been sequenced. Therefore, differential protein profiles could serve as biomarkers themselves, whereby different cell lines may be compared on a global level to assess changes in a disease state or to distinguish specific states during disease progression. In this work, expression proteomics is applied to a squamous cell carcinoma of the human tongue, which has been shown to have decreased adhesive properties compared to normal cells.[15,16]

3.1.2 Reduced Cell-Cell Adhesiveness in Cancer Cells

In the 1940s, researchers noticed that cancer cells have reduced cell-cell adhesiveness compared to that of the corresponding normal cells.[17,18] Disruption of normal cell-cell adhesion triggers the release of cancer cells from the primary cancer nests, followed by invasion and metastasis of a tumor.[19] Other than decreased adhesion, major changes in the organization of the cytoskeleton and irregular adhesion-mediated signaling are indicative of malignant transformations.[20]

A simplified diagram of the cell adhesion system at the adherens junction is shown in Figure 3.1.[20,21] The adhesion of cells to their neighbors is important for maintaining tissue architecture and cell polarity and regulating major cellular processes that are essential for motility, growth, differentiation, and survival.[20] Cell-cell adherens junctions (AJs), the most common type of intercellular adhesions, act to connect cells to each other and anchor the cytoskeleton to the plasma membrane. The formation of AJs involves homophilic interactions between E-cadherin (E-cad) molecules, the major cadherin in epithelial cells.[21] The extracellular domains of the calcium-dependent cadherin receptors on the surface of one cell selectively binds to another from a neighboring cell. The cytoplasmic domains of cadherins can bind to two structurally similar proteins, β-catenin and plakoglobin (Pg, is identical to γ-catenin), which link to the actin cytoskeleton via α-catenin. The other type of epithelial adhering junctions are desmosomes. In order for desmosome to assembe, Pg is essential to mediate the interaction between the desmosomal cadherins and the cytokeratin intermediate filaments.[22]

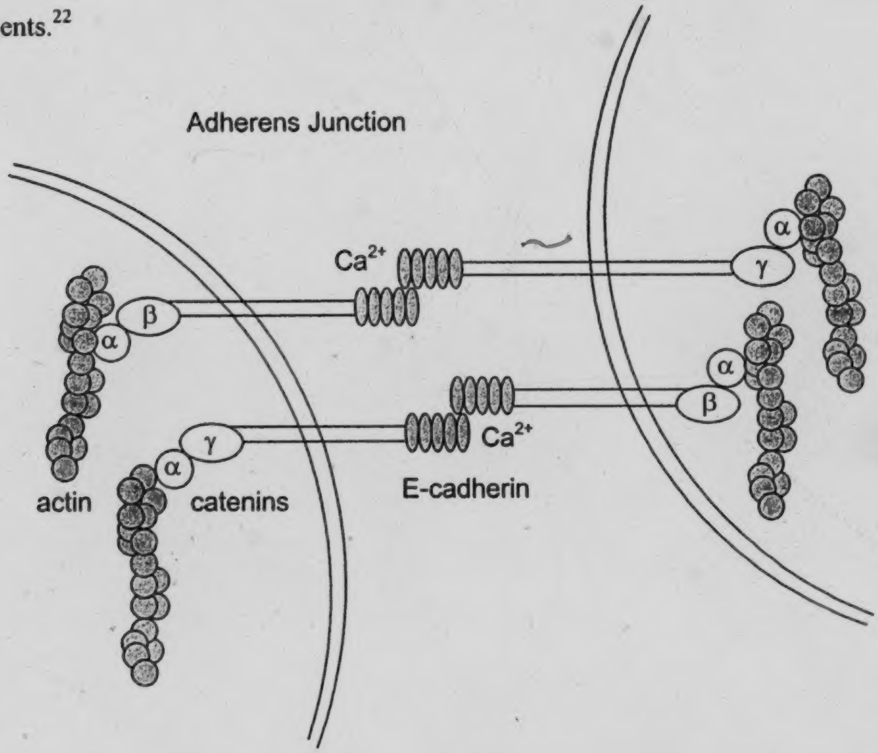


Figure 3.1 Diagram of the cell adhesion system at the adherens junction.

3.1.3 Mutations in Cell Adhesion Proteins

The disruption of cell-cell adhesiveness can be achieved by downregulating the expression of cadherin or catenin family members or by impedement of AJs assembly by activation of signaling pathways.[20] Reduced AJ assembly may occur by mutations, hypermethylation, and transcriptional repression of E-cad expression and may contribute to enhanced motility and invasiveness of cancer cells.[20] Cancer therapies targeting the regulation of E-cad expression may offer powerful means to tumor-suppressive effects.[23]

When the cadherin molecules become mutated, they lack the catenin-binding sites and fail to interact with α-catenins and therefore cannot link to the actin filaments. Loss of E-cad expression is commonly observed in many types of carcinoma, while E-cad reintroduction into cancer cells reduces invasion and metastasis.[21] When E-cad is sufficiently active, epithelial cells, including cancer cells, cannot disrupt their mutual connections, but suppression of E-cad activity may trigger the release of cancer cells from primary cancer nests.[21]

The abnormal expression of β-catenin and Pg also disrupt normal cell adhesion. Pg is a structural and functional homologue of β-catenin with dual adhesive and signaling roles for regulating cellular fate determination, growth and proliferation. Structurally, Pg and β-catenin proteins have similar binding sites for various cellular proteins including cadherins, the transcription factors of LEF/TCF family, and the tumor suppressor protein adenomatous polyposis coli (APC).[23] If not sequestered by APC, excess β-catenin does not get degraded and remain in the cytoplasm. β-catenin acquires oncogenic activity when it is mutated or when it is up-regulated as a result of inactivation of APC. In vitro experiments revealed that the direct fusion of α-catenin to the cytoplasmic tail of cadherin molecules produces adhesion-competent cells in the absence of β-catenin.[24] The role of β-catenin in the cadherin-catenin complex is believed to be regulatory, by bringing α-catenin into close association with cadherin.[25]

3.1.4 Squamous Carcinoma Cells

A squamous cell carcinoma line, SCC9, that lacks both Pg and E-cad and does not assemble desmosomes was used in this study.[15,16] These cells have both α- and β-catenins but express a significant amount of N-cadherin, which is not normally expressed in epithelial cells. The morphology of SCC9 cells is similar to that of transformed cells, consisting of heterogeneous populations of long (fibroblast-like) and rounded cells that grew together. Previous research on SCC9 cells were conducted to determine the role of cell adhesion proteins in the development of transformed phenotype. In one study, transfection of SCC9 cells with a plasmid containing cDNA encoding the functional homologue of E-cad, chicken L-CAM, induced a morphologic transformation from fibroblastoid to epidermiod.[15] The cells became flatter, formed tight epithelial monolayers and showed growth and adhesive properties typical of epidermoid morphology. The transition of 'transformed' to 'normal' phenotype coincides with downregulation of the endogenous N-cadherin and increased synthesis and stability of the catenins. In another study, it was shown that the introduction of a low amount of Pg cDNA into SCC9 cells enabled desmosome assembly and increased adhesivity and similarly, induced a fibroblast to epidermoid transition.[16] However, unlike the effects of induced E-cad expression, these Pg-induced changes coincide with increased stability and steady-state level of N-cadherin and decreased level and stability of β-catenin without any significant effects on α-catenin. When expressed at high levels, however, Pg induced unregulated growth and foci formation.[23] Thus, the amount of Pg and β-catenin and their interaction with cellular partners in the cell adhesion system are important and must be regulated.

The purpose of this initial work is to look at the changes in protein expression of SCC9 cells transfected with Pg or E-cad since these proteins were lacking in the transformed cells. The control cell line is referred to as SCC9, while the resulting cell line after introduction of E-cad and Pg expression are referred to as SCC9-E-cad and SCC9-Pg respectively. In this work, mass spectrometry combined with separation techniques was used for profiling of protein expression in SCC9, SCC9-E-cad and SCC9-Pg cells. The results of three different analysis methods were compared: 2D gel electrophoresis, direct MALDI MS and HPLC fractionation with offline MALDI MS. 2D gel electrophoresis was used to provide a map of the proteome, and visualize any differentially expressed proteins. Under the sample preparation employed, the separation of the extracted proteins displayed in the 2D gels is between 20-100 kDa mass range. The change in expression of the low mass proteins below 20 kDa were detected using direct MALDI MS and HPLC offline MALDI MS. The relatively simple sample preparation method employed, generally limited the observation of peaks in MALDI TOF MS to water-soluble proteins under 20 kDa, these small proteins may be important diagnostic biomarkers. Direct MALDI MS was used to preview the total cell extract in the cancer cells. HPLC offline MALDI MS was used to simplify the complex extract by reducing ion suppression effects and the low mass proteins were catalogued in a mass table. The study is used to obtain a better understanding of the cadherin-catenin system in the regulation of cell proliferation, invasion, and intracellular signalling during cancer.

3.2 Experimental

3.2.1 Materials and Reagents

Sodium dodecyl sulfate (SDS), α-Cyano-4-hydroxycinnamic acid (HCCA), ethylenediaminetetraacetic acid (EDTA), phenylmethanesulfonyl fluoride (PMSF), and horse cytochrome *c* were purchased from Sigma-Aldrich Canada (Markham, ON, Canada). HCCA was recrystallized from ethanol (95%) at 50 °C prior to use. Analytical grade acetone, methanol, acetonitrile, formic acid and trifluoroacetic acid (TFA) were from Caledon Laboratories (Edmonton, AB). HPLC grade acetonitrile from Fisher Scientific Canada (Edmonton, AB, Canada) was used in HPLC experiments. Water used in all experiments was from a NANOpure water system (Barnstead/Thermolyne). Bovine trypsin, glycerol, and 2-mercaptoethanol were obtained from Sigma (St. Louis, MO). Tris(hydroxymethyl)-aminomethane, acrylamide/bis (electrophoresis purity grade), and Coomassie Blue G250 staining solution were obtained from BioRad (Hercules, CA). The peptide extracts were desalted using commercial C18 μZipTips from Millipore (Bedford, MA). Protein extracts were desalted using 3 kDa cut-off Microcon filter (Bedford, MA).

3.2.2 Sample Preparation

The human squamous carcinoma cells SCC9 are derived from carcinomas of the tongue (American Type Culture Collection (ATTC), Rockville, MD) and provided by Dr. M. Pasdar (Department of Cell Biology and Anatomy, University of Alberta, Edmonton, Alberta, Canada). The cell lines were maintained in minimum essential medium (MEM) (Sigma, St. Louis, MO) supplemented with 10% fetal bovine serum (FBS).[16] The cells were prepared by Dr. Laiji Li of Professor Manijeh Pasdar's research group at the Department of Cell Biology and Anatomy, University of Alberta, Edmonton, Alberta, Canada. SCC9 cells were grown to about 80% confluency in 150 mm tissue culture dishes. The cells were washed three times with 150 mM sorbitol in 10 mM phosphate buffer (pH 7.0) prior to scraping with a rubber policeman and the cell pellets were stored in aliquots under - 80 °C.

The Pasdar's group performed two transformations of the SCC9 cells in this study and reported elsewhere. SCC9 cells transfected with E-cad cDNA[15] are referred to as SCC9-E-cad cell line in this report. The SCC9-E-cad cells show decreased N-cadherin and increased catenin expression (compared to the control cell line SCC9). SCC9 cells transfected with Pg cDNA[16] are referred to as SCC9-Pg cells. The SCC9-Pg cells show increased N-cadherin and decreased β-catenin expression, the opposite of that observed for SCC9-E-cad cells. Table 3.1 summarizes the transformation performed and the cell adhesion proteins expressed in these cells.

The sample preparation of the squamous carcinoma cells for suitability for 2D gel electrophoresis, direct MALDI MS, and RP HPLC offline MALDI MS analysis are outlined below. Ultracentrifugation of solubilized biological membranes (at 100,000 g, for 15-60 min) is an essential step that removes insoluble components that would impede further purification steps. EDTA and PMSF must be used with care since the former may bind to other chromatography media, while the latter is insoluble in water. Removal of these contaminants by ultrafiltration must be performed prior to electrophoresis or chromatography. Protease inhibitors, such as EDTA and PMSF, were employed in our experiments and no such problems associated with proteolysis were observed in the gel or in the chromatographic separations of the extracted proteins.

Table 3.1 The cell adhesion proteins expressed in SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue.

	Squamous Carcinoma Cell Lines		
Expression	SCC9 (control)	SCC9-E-cad	SCC9-Pg
Plakoglobin (Pg)	×	×	√ (transfected with Pg cDNA)
E-cadherin (E-cad)	×	√ (transfected with E-cad cDNA)	
N-cadherin	√	↓	↑
β-catenin			↓

3.2.3 2D-Gel Electrophoresis

For all 2D gel electrophoresis experiments, SCC9, SCC9-E-cad and SCC9-Pg cells were treated with DNAse and RNAse to remove nucleic acids that could cause background smears in the gel. An overnight acetone precipitation was performed to pellet the proteins and get rid of the soluble contaminants, such as detergents, lipids and small ionic molecules, present in typical cell extracts. After IEF, equilibration steps with DTT and iodoacetamide was performed prior to loading the strips onto the second dimension.

3.2.3.1 Reproducibility Study of 2D-PAGE Method

The reproducibility of the 2D gels was tested, two gels for each of the three samples were run in parallel mini gel format, with the same running conditions. The IPG strips with pH range 3 to 10 and 7cm length were used. The gels were cast as 10% separating gels with 1 mm thickness. The determination of protein concentrations (protein assay method using BSA standard for 2D gels) extracted from the cells were 2 mg/mL for SCC9, 2.55 mg/mL for SCC9-E-cad and 2.1 mg/mL for SCC9-Pg. 100 μg of protein was loaded for each gel to maximize the amount of sample loaded without overloading.

IEF Running Conditions:

Step 1 - 100 V for 1.5 hr

Step 2 - 500 V for 30 min

Step 3 - 1000 V for 30 min

Step 4 - 5000 V for 15,000 Vhr

Step 5 - 500 V, stop

SDS PAGE Running Conditions (for running 6 gels at the same time):

180 V constant for 45 min

3.2.3.2 Protein Profiling by 2D-PAGE

The final 2D gels were obtained for SCC9, SCC9-E-cad, and SCC9-Pg using a large get format. The IPG strips with pH range 3 to 10 and 13cm length were used. The gels (16X20 cm) were cast as 12% separating gels with 1 mm thickness. Concentrations were determined as 1.52 mg/mL for SCC9, 1.63 mg/mL for SCC9-E-cad and 1.71 mg/mL for SCC9-Pg. 400 μg of protein of each sample was loaded onto the 13 cm IPG strips.

IEF Running Conditions:

Step 1 - 100 V for 1.5 hr

Step 2 - 500 V for 1 hr

Step 3 - 1000 V for 1 hr

Step 4 - 8000 V for 24,000 Vhr

Step 5 - 500 V, stop

SDS PAGE Running Conditions: 25 mA constant for 4.5 hr

3.2.4 Direct MALDI MS and RP HPLC Offline MALDI MS

Summary of Experimental Procedure:

I. The pelleted cells were suspended in water containing protease inhibitors (1mM EDTA and 1mM PMSF) and cell disruption was performed with the French Press, proteins were extracted with 0.1% TFA, ultracentrifuged at 100,000 g for 30 min and the supernatant was collected.

II. The extract mixture was speed-vac concentrated by four times and desalted using 3 kDa cut-off Microcon filter. Protein concentrations were determined by the method of Bradford[26] using a Coomassie Plus-200 dye reagent and an IgG standard (Pierce Chemical Co.).

III. Direct MALDI MS (1 μg in 10 μL HCCA solution) was used to preview the proteins in the total cell extract.

IV. HPLC offline MALDI MS (~50 μg protein loaded) was used to separate proteins into 1 min fractions and detect the proteins in the extract mixture by eliminating effects from ion suppression.

V. The low mass proteins (<20 kDa) were catalogued in a mass table.

RP HPLC offline MALDI MS:

I. Solvent delivery and separations were performed on an Agilent (Palo Alto, CA) HP1100 HPLC system. The solvents used for reversed-phase HPLC separation were water (A) and acetonitrile (B) with 0.1% (v/v) TFA in both phases. The gradient elution profile was optimized for the control cell line, SCC9 extract.

II. For LC offline MALDI MS, ~50 μg of total protein extract was separated on a 150 × 1.0 mm i.d. C8 column (5 μm particles with 300 °A pore size; Vydac, Hesperia, CA) at a flow rate of 40 μL/min. Fractions were collected at 1-min intervals using an Advantec SF-2120 super fraction collector.

III. Fractions were concentrated by 8 times to ~5 μL before mixing with matrix for analysis. A two-layer method was used for MALDI MS sample preparation. An aliquot of the fraction was diluted 1:10 with the second layer matrix for optimal ionization for peak detection in MS. HCCA was used as the matrix. About 0.5 μL of the first layer (16 mg/mL HCCA in 26%MeOH/acetone) was deposited on the MALDI probe. About 0.5 μL of the second layer (saturated HCCA in 1:2:3 (v/v/v) Formic acid:MeOH:water) with the sample solution was deposited and air-dried.

IV. On-probe washing of the MALDI sample spot with water was performed twice to remove the salts. External mass calibration was done in the mass range of 4-12 kDa using horse cytochrome *c* and its multiply charged species and dimer. With external calibration, the mass measurement accuracy is normally better than 0.05%.

The mass values of singly protonated molecular ions ($[M+H]^+$) are listed in Tables 3.3, 3.4, 3.5, and 3.6 for the direct MALDI MS and RP HPLC offline MALDI MS experiments. Clusters of peaks with a mass difference of 16 Da in the molecular ion region of the protein were sometimes observed. These peaks correspond to methionine-oxidized (Met-ox) proteins which most likely occurred during protein extraction or MALDI sample preparation. For the mass tables listed in this work, only the non-oxidized protein masses were included.

3.3 Results and Discussion

Three different approaches were performed to investigate and compare the differences in the proteins observed for SCC9, SCC9-E-cad, and SCC9-Pg cells. Often the detection of proteins with molecular masses greater than 100 kDa and those with pI values less than 4 or higher than 9 is difficult to detect using 2D gels.[7] In this work, the use of 2D-gel electrophoresis generated a proteomic map of the extracted proteins between 20-100 kDa mass range while the methods of direct MALDI MS and RP HPLC offline MALDI MS generated proteins in the mass range below 20 kDa. The proteomic profiling of these transformed cells has not been previously investigated in the literature.

3.3.1 Protein Profiling by 2D PAGE

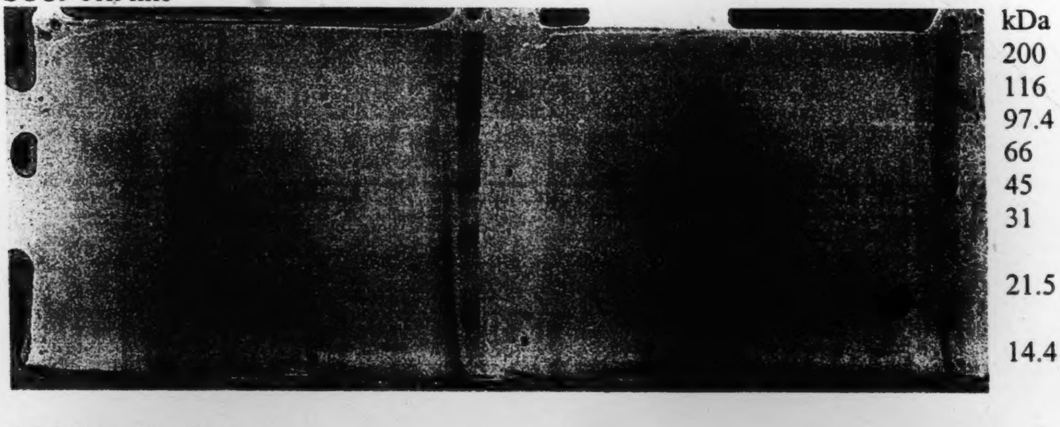
3.3.1.1 Reproducibility of the 2D PAGE Method

The method of 2D gel electrophoresis employed exhibited difficulty in the separation of proteins that are in the very low and very high MW range, as observed in Figure 3.2 and 3.3. The 10% separating gel used in casting the mini gels instead of the typical 12% normally used was to improve the detection of proteins with large MW especially since mini gels were used. The proteome map generated by 2D gel electrophoresis provides both intra-gel (giving relative abundance data within a cell) and inter-gel (giving quantitative variations between different conditions) quantitative data. A reproducibility study was conducted in which two 2D mini gels were run per cell type to verify that the method employed resulted in reproducible gels and therefore allows for a reliable comparison between the SCC9, SCC9-E-cad, and SCC9-Pg cells.

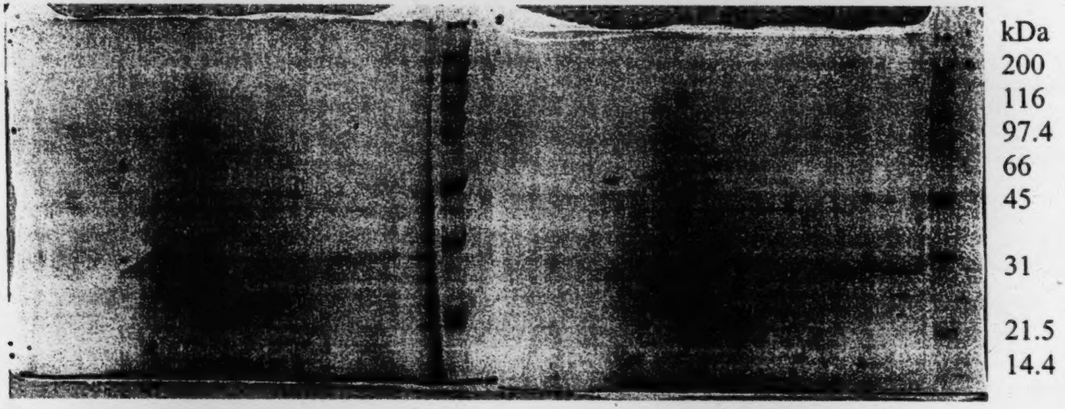
The comparison of the two gels of a particular type of cell, in Figure 3.2, showed that there was little if any intra sample variation and therefore a reliable inter-gel comparison can be studied. Differences in sample handling, such as from sample loss during the purification steps or loading of sample onto the gel can result in the differences observed in the two gels produced from each cell type. For SCC9 cells, the two gels look different in the relative spot intensity, likely due to sample loss during the desalting step in one gel and not the other. Problems in loading of the sample onto the second dimension gel caused horizontal streaking in one gel of SCC9-E-cad cells.

There are no significant spot differences between the different cells, in other words, no significant inter-sample variations. As shown in Figure 3.2, there are not a lot of observable differences between SCC9, SCC9-E-cad, and SCC9-Pg cells in the resultant 2D mini gels since the spots are visualized in such a small area of the mini gel. The most intense spots are similar in all gels. Many spots are faint due to the low amount of protein, typically loaded onto mini gels. In the next experiment, a larger gel format was used to increase the protein load, to be able to observe differences in not so abundant proteins.

SCC9 cell line


kDa
200
116
97.4
66
45
31
21.5
14.4

SCC9-E-cad cell line


kDa
200
116
97.4
66
45
31
21.5
14.4

SCC9-Pg cell line

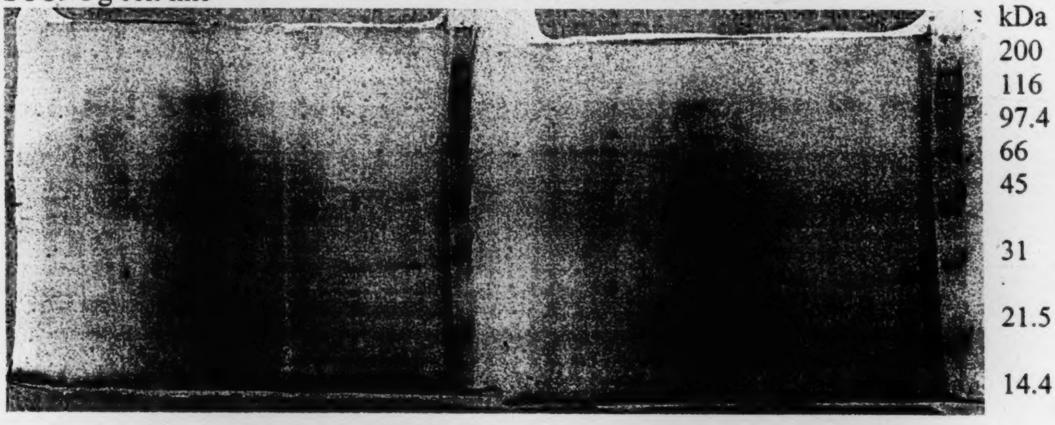

kDa
200
116
97.4
66
45
31
21.5
14.4

Figure 3.2 Reproducibility study of 2D gel electrophoresis of Human squamous carcinoma cells of the tongue. 100 μg total protein was loaded.

3.3.1.2 Proteome Map Generated by 2D PAGE

The resultant large 2D gels of SCC9, SCC9-E-cad, and SCC9-Pg cells are shown in Figure 3.3, 3.4, and 3.5 respectively. The 2D gel of the SCC9 cells showed uneven polymerization of the second dimension gel. The horizontal streaking may be due to overloading of the sample or poor focusing of the spots. Note that the gel for SCC9-Pg broke during destaining because the gels are very fragile (only 1mm thick, 16X20 cm gels). However, the sample handling variation listed above does not hinder the inter-gel comparison. An inter-gel comparison of protein expression in SCC9, SCC9-E-cad, and SCC9-Pg cells indicate some differences, in particular, the quantity of a specific spot in one cell type is different in the other cells. These spots, numbered as 1, 2, 3, 4 in SCC9 cells, 5, 6, 7, 8 in SCC9-E-cad cells, and 9, 10, 11, 12 in SCC9-Pg cells, show distinct changes between the cell types and were investigated further by performing trypsin digestion and subsequent peptide mass mapping by MALDI mass spectrometry and identification by MS/MS. Spot 4 in SCC9 cells show separated isoforms, whereas, the corresponding spot with similar pI and molecular mass coordinate in the other gels, spot 8 in SCC9-E-cad cells and spot 12 in SCC9-Pg cells, just show one form. The poor MS/MS spectra obtained, likely from the low amount of protein available and difficulty in fragmentation of the selected peptides, inhibited the identification of spots 4, 8 and 12. The identification results from the spots are summarized in Table 3.2. The biological significance of these identified proteins needs to be investigated in the future.

SCC9 cell line - 400 μg total protein

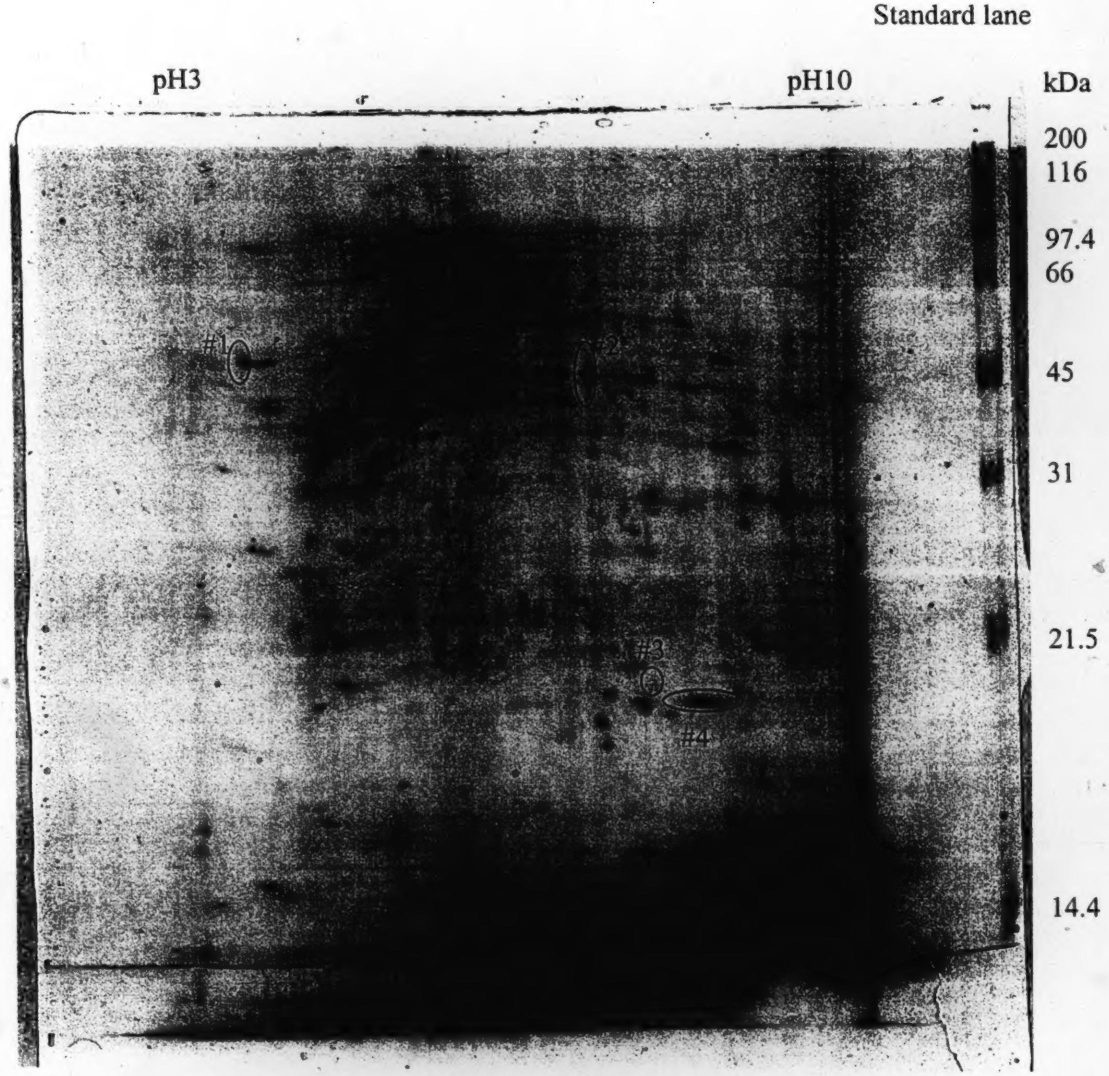

Standard lane
pH3
pH10
kDa
200
116
97.4
66
45
31
21.5
14.4
#1
#2
#3
#4

Figure 3.3 Protein profiling of SCC9 cells of the Human squamous carcinoma of the tongue by 2D gel electrophoresis MALDI MS.

SCC9-E-cad cell line - 400 µg total protein

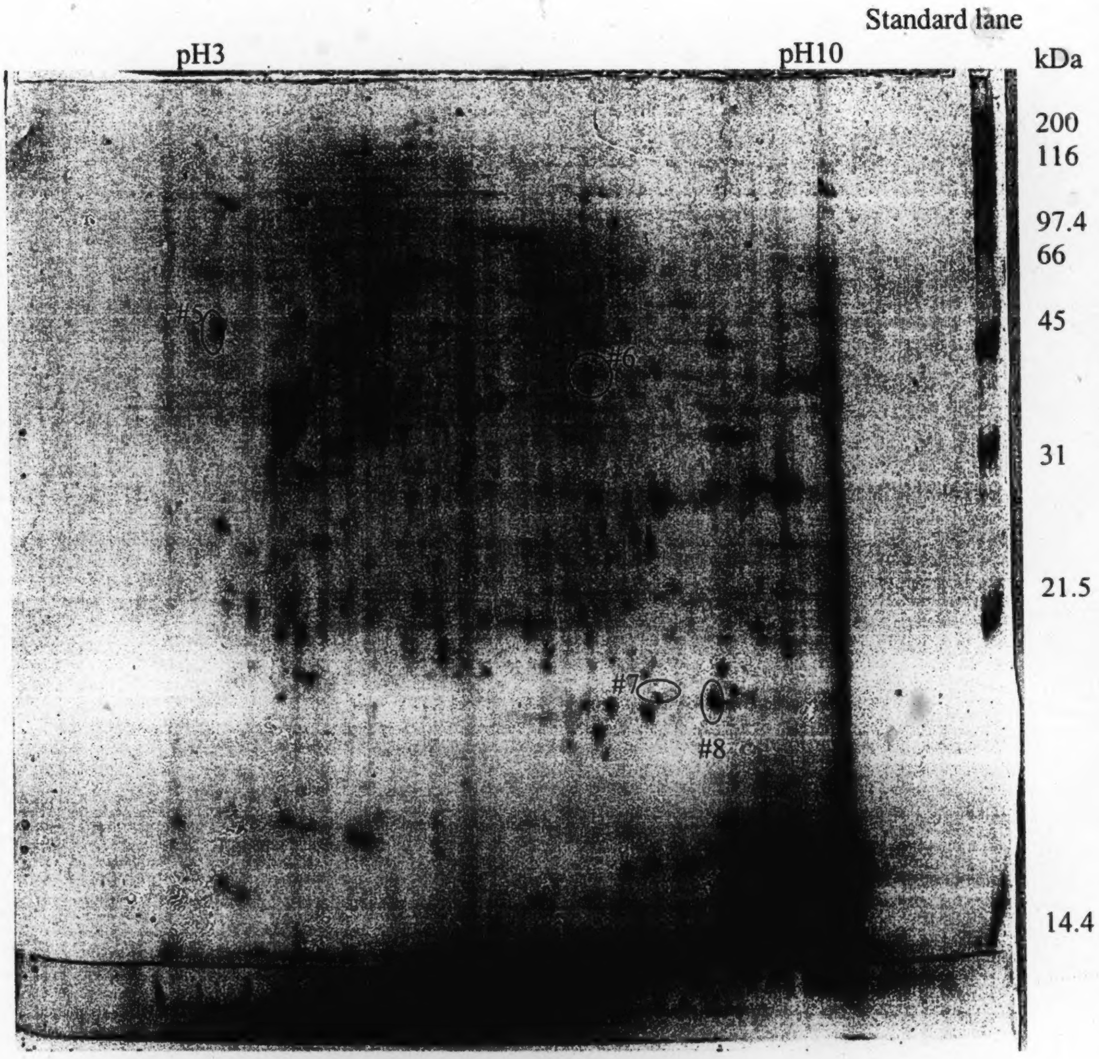

Standard lane
pH3
pH10
kDa
200
116
97.4
66
45
31
21.5
14.4
#5
#6
#7
#8

Figure 3.4 Protein profiling of SCC9-E-cad cells of the Human squamous carcinoma of the tongue by 2D gel electrophoresis MALDI MS.

SCC9-Pg cell line – 400 µg total protein

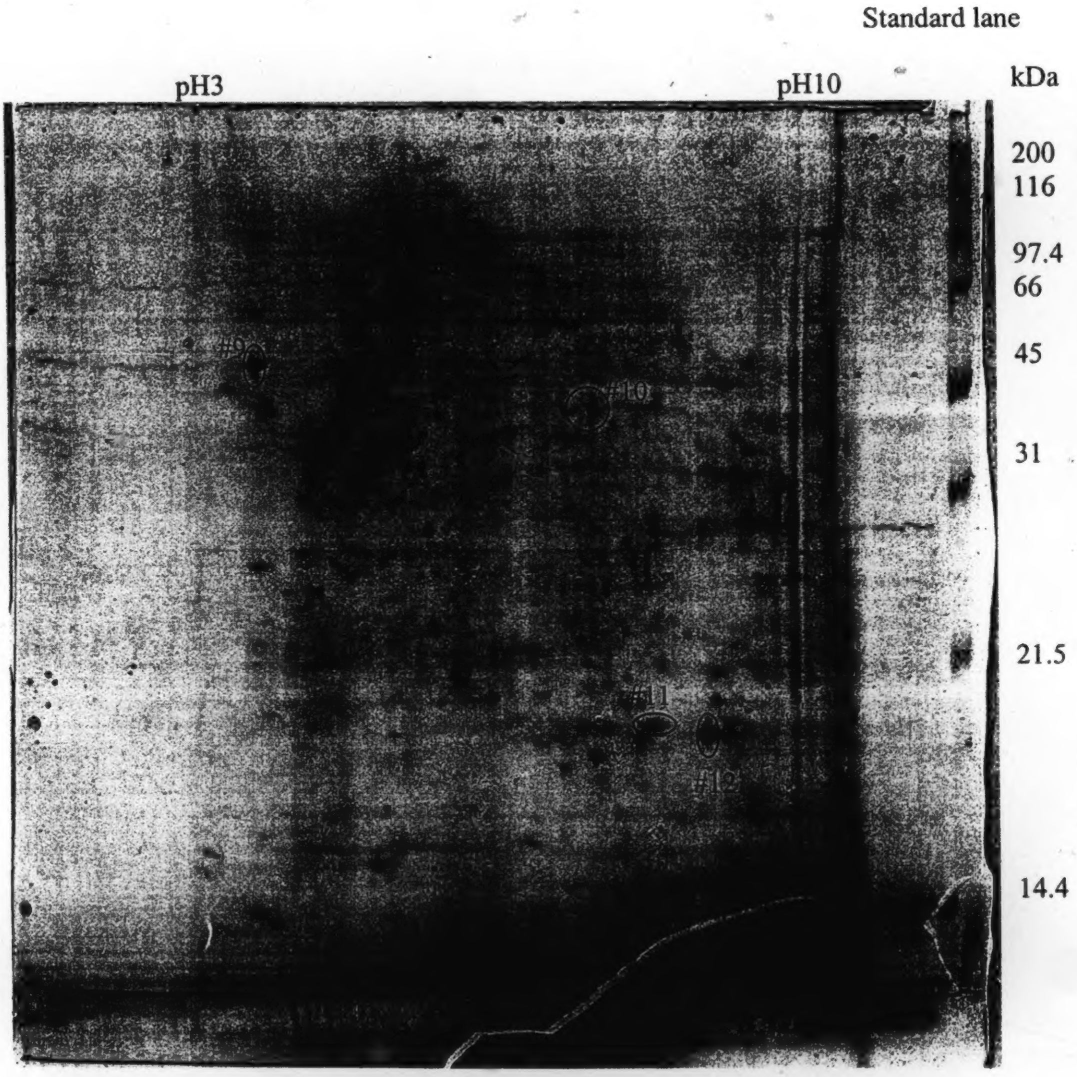

Standard lane
pH3
pH10
kDa
200
116
97.4
66
45
31
21.5
14.4
#9
#10
#11
#12

Figure 3.5 Protein profiling of SCC9-Pg cells of the Human squamous carcinoma of the tongue by 2D gel electrophoresis MALDI MS.

Table 3.2 Identification of proteins showing distinct changes in SCC9, SCC9-E-cad, and SCC9-Pg cells of the Human squamous carcinoma of the tongue by trypsin digestion, MALDI TOF MS and MS/MS analysis.

Protein Identity	Cell Type		
	SCC9	SCC9-E-cad	SCC9-Pg
Calreticulin precursor, pI 4.3, MW 48 kDa	Spot #1	Spot #5	Spot #9
Alpha enolase, pI 7.0, MW 47 kDa	Spot #2	Spot #6	Spot #10
Peroxiredoxin (proliferation-associated gene A), pI 8.3, MW 22 kDa	Spot #3	Spot #7	Spot #11
Unidentified	Spot#4	Spot #8	Spot #12

The resultant 2D gels of protein extracts of SCC9, SCC9-E-cad, and SCC9-Pg cells display the proteome map in a concise manner and allow for a direct comparison of the expressed proteins. However, observing differences in protein expression on a 2D gel is limited by the close separation of many protein spots having similar pI and MW values onto a small area of the gel. Improvement in the procedure is necessary to simplify the complex mixture using the 2D gel approach, for example, using narrow pH range IPG strips can be used to zoom in and expand the gel image for better separation. With each additional step, sample loss can be critical. The preliminary work in profiling of the proteins in SCC9, SCC9-E-cad, and SCC9-Pg cells by the method of 2D gel electrophoresis visualized proteins between 20 to 100 kDa proved to be quite labour intensive.

2D gel electrophoresis followed by protein identification by mass spectrometry has been the conventional method in proteome analysis. The direct comparisons of protein expression and identification of the differentially expressed proteins between normal and tumour tissues have been well suited for biomarker discovery in various cancers, such as liver, bladder, lung, esophageal, prostate and breast.[6] Limitations of the 2D gel approach are the large amount of starting protein required, the unreliable detection and identification of low abundance proteins and low throughput, requiring days for analysis. The development of integrated systems that use robots to automatically pick specified spots on a two-dimensional gel, perform enzymatic digestion, and process the samples for analysis through MALDI TOF MS offer a more rapid approach, however extensive effort of the investigator is still required. The method used for profiling of protein expression in the lower mass range, <20 kDa, is similar to previous work by Wang et al[27] applied towards bacteria differentiation through the detection of unique 'biomarkers' or organism-specific ions. In the following sections, the methods of direct MALDI MS analysis of the extract solution and performing HPLC to fractionate the complex extract mixture to simplify the MALDI mass spectra were compared.

3.3.2 Protein Profiling by Direct MALDI MS and RP HPLC Offline MALDI MS

The extraction from tissue with 0.1% TFA was used to selectively precipitate large proteins while enhancing the solubility of smaller proteins and peptides due to their ion-pairing properties.[28] TFA is volatile and compatible with subsequent fractionation by reversed phase HPLC. Acid cleavage of proteins containing Asp-Pro sequences may contribute to the population of lower MW peptides in the extracts.[29] It is possible that we observed fragments of a larger protein due to acid cleavage, which complicates the extract mixture. The one-dimensional gel (not shown) of the TFA extract of the squamous carcinoma cells of the tongue show that the proteins present in the extract also include proteins of high MW range, above 20 kDa. However, the reliable detection of proteins in the MW range below 20 kDa in the extract is observed with the technique of MALDI MS. The potential biomarker proteins detected in the MW range below 20 kDa could also be a fragment peptide of a larger protein.

3.3.2.1 Direct MALDI MS

While the direct MALDI MS (1 μg in 10 μL HCCA solution) is used to preview the proteins in the total cell extract, the HPLC offline MALDI MS (~50 μg protein loaded) is used to separate proteins into 1 min fractions and detect the proteins in the extract mixture by minimizing effects from ion suppression. In this comparative study, the water-soluble low mass proteins (<20 kDa) are detected in the mass spectra, labelled and then catalogued in mass tables, with the potential identified protein in the extract being represented by their m/z value. The multiply charged species were not included in the mass tables, but only peaks corresponding to singly charged species having higher than signal-to-noise (S/N) ratio of 3 were labelled. As mentioned in the experimental, an external mass calibration was done in the mass range of 4-12 kDa using horse cytochrome *c* and its multiply charged species and dimer. With external calibration, the mass measurement accuracy is normally better than 0.05%. The mass specta of the direct MALDI MS experiment is shown in Figure 3.6 and the data is summarized in Table 3.3, which can be compared with that obtained using the HPLC offline MS in Table 3.6. In the mass tables, the matched masses are placed along the same row for each cell line to indicate that they may have potentially the same identity. However, the same m/z value detected in several extracts and/or fractions may not have the same identity. In this preliminary work, peaks having similar m/z values are recognized as potentially representing the same identity.

The total number of proteins observed using direct MALDI MS, summarized in Table 3.3, are 8 proteins for SCC9, 6 proteins for SCC9-E-cad, and 12 proteins for SCC9-Pg cell lines. LC fractionation used to simplify the complex extract mixture prior to MS analysis is clearly an advantageous step in protein profiling. The resulting mass spectra are much improved in the S/N ratio, resolution and mass accuracy and also more peaks are detected in the extract using HPLC offline MALDI MS. The fractionation reduces the number of proteins present in the resulting MALDI sample spot, minimizes effects of ion suppression from more ionisable proteins and allows for a higher probability for each protein to be detected. In the direct MALDI MS experiment, the variation in concentration of the proteins means that the signal from the less abundant proteins will be insignificant compared to the high abundant proteins in the total cell extract (a dynamic range issue), which have led to the higher number of detected species when HPLC offline MALDI MS was used. The HPLC chromatograms of the squamous carcinoma cells are shown in Figure 3.7.

Although direct MALDI MS analysis is a straightforward method, it clearly suffers from the inability to detect a large number of components present in the cell extracts. In contrast, LC/offline MALDI MS analysis can detect 5-10 components in one particular fraction. For example, compare the mass spectra of the direct MALDI MS in Figure 3.6 to one fraction, the fraction at time 24 min, of the HPLC offline MALDI MS experiment, shown in Figure 3.8. The fractionation helped to detect many more components in the cell extract in the MALDI MS by reducing ion suppression effects. Using one-dimensional HPLC separation, each fraction consists of many protein components. Thus, ion suppression still remains an obstacle, although not to the same extent as with direct MALDI MS. Another reason for a protein not to be detected by HPLC offline MALDI MS is that some proteins could have eluted into several adjacent HPLC fractions and become too dilute to be detected after HPLC fractionation. The potential for protein loss during sample workup, such as protein concentration by solvent evaporation in individual fractions, can also contribute to the differences observed.

Table 3.3 Protein profiling of SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue by direct MALDI MS. ($[M+H]^+$) refer to the singly protonated molecular ions detected and are listed by their m/z values. *This is the oxidized component of ~6513.

SCC9	SCC9-E-cad	SCC9-Pg
$[M+H]^+$	$[M+H]^+$	$[M+H]^+$
		2979
	3456	3456
		3889
		4631
4938	4937	4937
4965		
		5357
6277		6276
6513	6512	6527ox*
6649	6647	6647
		8450
9262		
11301	11303	11308
13770	13769	13770

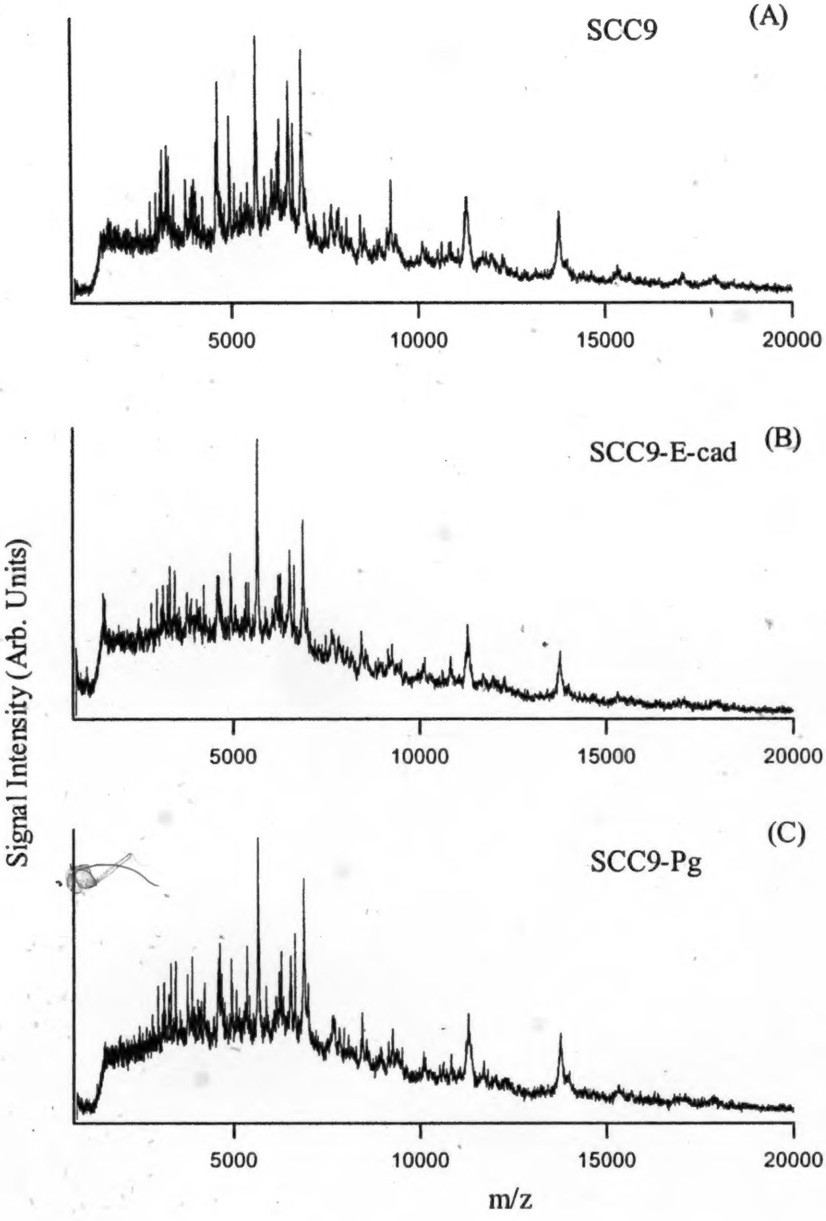


Figure 3.6 Mass spectra of direct MALDI MS of Human squamous carcinoma cells.

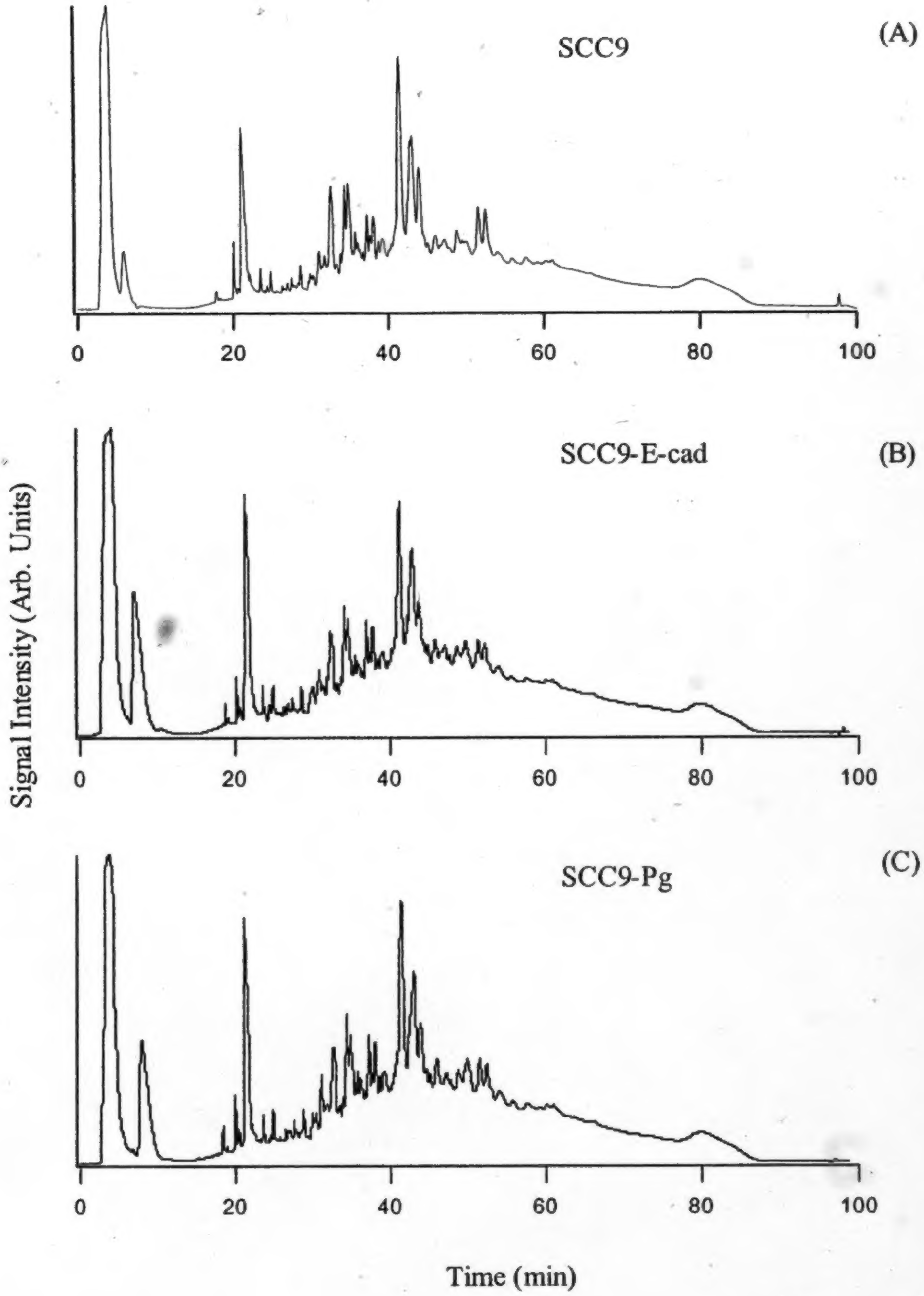


Figure 3.7 HPLC chromatograms of Human squamous carcinoma cells.

3.3.2.2 Reproducibility Study of the HPLC Offline MALDI MS Method

The HPLC MALDI MS data collected three times for the same extract of SCC9 cells (Table 3.4) was overall reproducible indicating that the separation and collection/detection of the proteins using this method was successful for this complex extract mixture. The data represented is more reproducible in the lower mass range, < 10,000 Da. Although, less reproducibility is observed in the high mass range, the data can still be credited as reproducible since problems in the detection of high mass proteins is inevitable in MS. The peaks in the mass spectra in this range are generally broader and give rise to poorer mass accuracy. The problem in mass accuracy also makes it difficult to judge whether the peak is the oxidized peak since the acceptable range of 15-17 Da and 31-33 Da can be increased when the error is large. Oxidation will limit the detection of real peaks and complicate the analysis. More rigorous sample preparations to prevent oxidations during sample work up should be investigated.

The decrease in reproducibility at the high mass range may be due to the difficulty of detecting larger proteins and the associated problem with mass accuracy when the peak is broad with low S/N. This is also in agreement that detection sensitivity is worse when MW increases. Therefore, the reproducibility of the extraction method of the cells is dependent on the mass range of the proteins. However, the reasons for decreased reproducibility in the high mass end of the mass spectra can also be attributed to the difficulty of extracting larger proteins due to precipitation problems and sample loss. In this experiment, the same extract is run three times using HPLC offline MALDI MS, all proteins in the extract should be detected in all of the repeated runs. This was not the case and therefore, the error lies in the difficulty in the detection of high masses in MS and/or differences in sample recovery after HPLC and differences in ionization during MS.

Table 3.4 Reproducibility study of RP HPLC offline MALDI MS of the Human squamous carcinoma cells of the tongue, three repeated runs of the same extract mixture of SCC9 cells. The component with potential oxidation is labeled with ox.

Trial #1	Trial #2	Trial #3
$[M+H]^+$	$[M+H]^+$	$[M+H]^+$
		2735
2979	2979	2979
	3832	
3889	3890	3889
	3947	
	3991	3992
		4052
4140	4139	4139
4792	4767	**4788**
4824	4808	4806
4938	4937	4937
4964	4964	4964
5023		5021
5355	**5354**	**5354**
5549	5549	5549
		5752
	5992	
6154		
6219	**6224**	**6221**
6276	**6273**	6274
6295		
6327	**6329**	6327
6363	6361	6364
6386		6388
	6403	6400
6447		
6513	**6514**	**6514**
6649	**6647**	6648
6722	6720	6722
6763		6762
6920		
6995		
		7044

7176		7174
7343	7347	
7445		
7704	7706	7708
	7728	
	7800	7804
7881	**7885**	7885
	7932	
8407	8405	
8449	8450	8452
8563		8566
8603		
		8860
8959	8957	
9153	9151	9153
9261	**9262**	**9264**
		9318
9333	9338	
		9370
9412	**9416**	**9413**
9515	9518	9514
9567		9572
9662	9664	9662
9979		
10043	10043	**10044**
10097	10100	10103
10139	**10144**	10145
	10228	**10234**
10274	10272	
10347	10346	10348
10443		10443
10517		10522
10612		10615
10742	10738	**10749**
	10750	
	10812	
10829	**10831**	**10834**
	10839	
	10940	
10953	10946	
11073	**11072**	11075
11154	11156	11154
11305	**11302**	**11307**
	11313	
11347	**11345**	**11346**

	11387	
11406	**11406**	11407
11649	11645	**11653**
11731		11732
11752	**11758**	**11769**
11832	**11839**	**11830**
11899		
	11952	
11983	11983	**11987**
12067	12066	
12112	12111	
12168	12166	**12173**
12322	12332	12321
12445	12454	12452
12646		
		12657
12882		
13157	13161	13163
13432	13431	13434
13532		13536
	13640	
	13786	13770
13776		13787
13819		**13816**
13848		13848
13973	13974	13971
14002	**14002**	**14007**
		14024
14044	14038	14045
14160	14163	14168
14300	14299	14302
	14354	
14332		
14422	14420	
14494	14492	14498
	14762	
		14707
14922		14938
	15050	
		15059
15186		
		15205
15308		
15337	**15344**	**15348**
15380		

15419		15418
15466	15464	15472
15478		
15495		15490
15665	15668	**15684ox**
15722		
	15694	
		15748
		15835
	15924	
15934		15938
16144	16148	16147
		16316
	16348	
16796		
		16813
16824	**16826**	
		16836
16955	16947	16962
17072		
	17097	17097
17105		17105
	17144	
17185		
17257	17263	17262
17398		
17629	17631	17636
	17642	
17774	17784	17781
17875	**17873**	17882
17926	**17924**	
		17989
18020	18029	18032
18060		
18330		18346ox
18421		18423
18427		
	18470	
18742		
18959	18950	18966ox
		19633
19701	19692	19701
	19892	

3.3.2.3 Reproducibility Study of Extraction Method

In the reproducibility study of three different extractions of SCC9 cells and the analysis by HPLC offline MALDI MS (Table 3.5), a similar set of proteins were detected in each extraction. However, more similarities were observed for masses < 10,000 Da (as observed in Table 3.4), meaning that the reproducibility of the experiment is better in this range. For proteins with higher mass (mainly >10,000 Da), reproducibility is not as good, as discussed above.

The times of elution of a particular mass that matched in all three different extractions of SCC9 cells is very reproducible. For example, in Figure 3.8, the component at m/z 6649 is observed in all three different extractions of SCC9 cells at times 24 and 25 minutes. Peaks with similar m/z values appear in several fractions that are not adjacent to each other are frequently observed due to retention differences during separation by HPLC of the particular protein (this is recorded as the averaged value in one trial indicated by bolding the masses). The component at m/z 10,836 is observed in several fractions, at times 21, 35, and 36 minutes (refer to Figure 3.9). A small portion of the protein with m/z ~10836 is first eluted off at 21 minutes compared to other components in the fraction. At 36 minutes, the peak with m/z ~10836 is the major component in the fraction. This is commonly observed in HPLC separations involving complex mixtures in which the more highly retained component may elute in several fractions before all is eluted off the column. This may also hinder the detection of the low abundance proteins if its components are divided in multiple fractions.

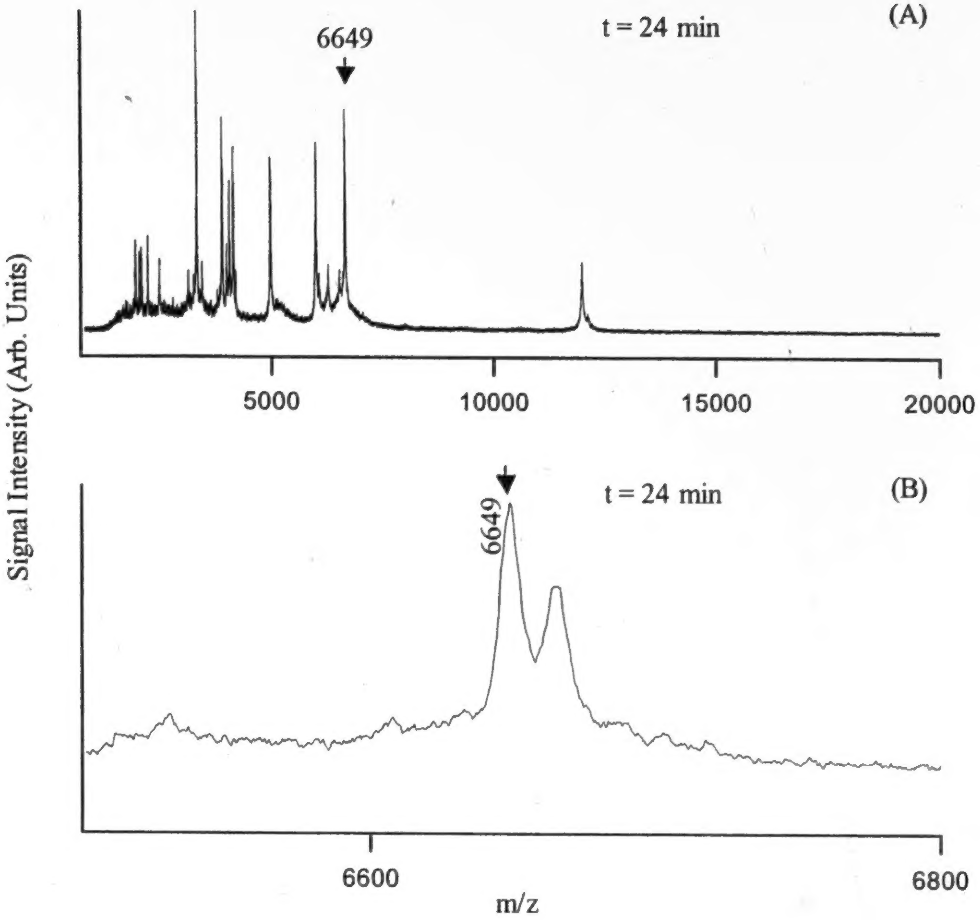


Figure 3.8 (A) Mass spectra of an HPLC offline MALDI MS experiment of a fraction at time 24 minutes. (B) The component at m/z 6649 is observed in all three different extractions of SCC9 at times 24 and 25 minutes.

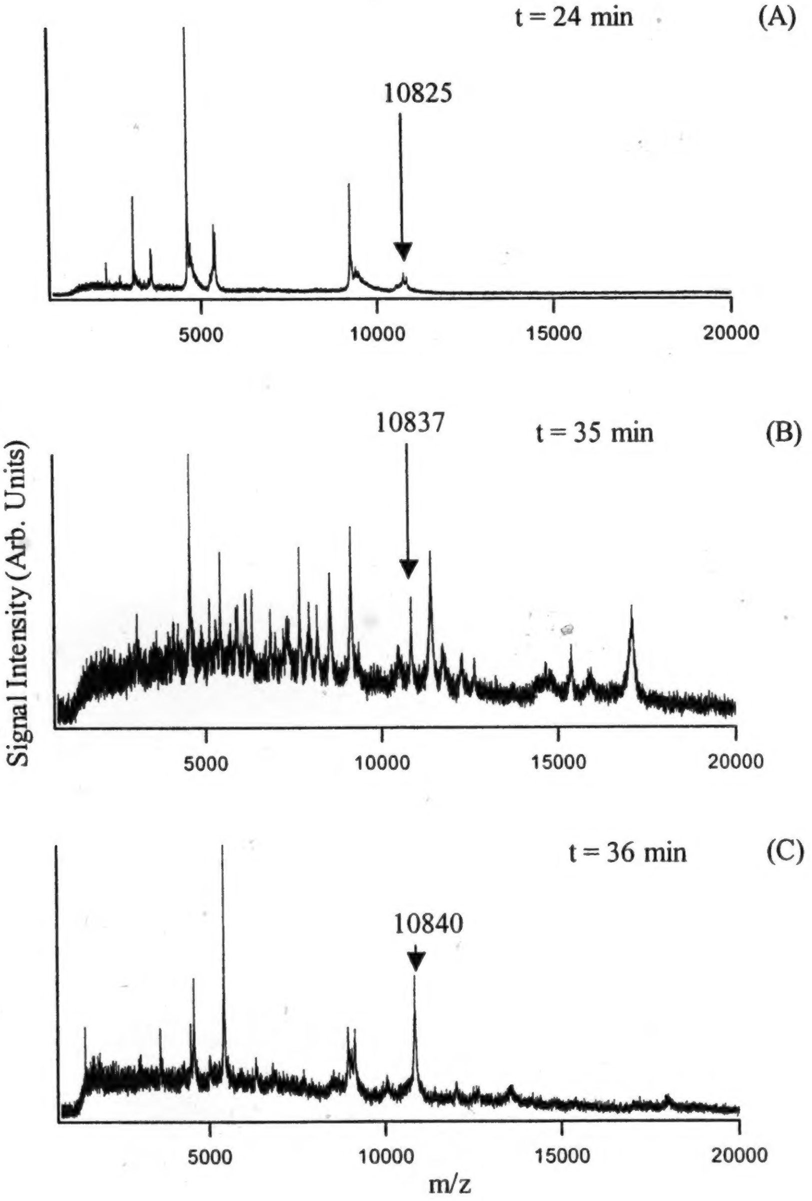


Figure 3.9 Mass spectra of HPLC offline MALDI MS of SCC9 cells. The component at m/z 10,836 is observed in several fractions, at times 21, 35, and 36 minutes.

Table 3.5 Reproducibility study of RP HPLC offline MALDI MS of the Human squamous carcinoma cells of the tongue, three runs of three different extract mixture of SCC9 cells. The component with potential oxidation is labeled with ox.

Trial #1	Trial #2	Trial #3
$[M+H]^+$	$[M+H]^+$	$[M+H]^+$
2980	2979	
3890	3906ox	3889
		3990
4053	4052	
4140	4140	4139
4793	4794	4792
	4808	4807
4939	4938	
4966	4981ox	4964
		5023
5357	**5357**	**5355**
5550	5550	
6070		
6222	**6222**	**6220**
		6232
6275	6276	
	6298	6300
	6330	6329
	6347	6349
	6365	6363
6400		6400
		6448
6459	6457	
6512	**6513**	**6514**
6581		
6649	**6650**	**6649**
	6690	
		6722
		6779
		6824
		7647
	7704	
	7802	
7883	7882	7883
8181		8181
8239	8238	8239
		8293
8404	8409	

8448	**8449**	**8449**
8563	8563	8564
8959	8957	8958
9151	**9150**	9152
9261	9262	**9262**
9313		
		9367
9413	**9412**	**9412**
9514	9517	9517
9533		
9568	**9565**	9569
	10098	**10095**
10109		
10200		
	10143	
		10274
	10348	10347
10518		10519
	10525	
10613	10607	
		10623
10748	10750	10752
10836	**10836**	**10838**
10896		
	11096	
		11110
11152		11156
11298		**11299**
	11307	
11652	**11655**	11650
		11730
11751	11747	11754
	11761	
11832		11830
	11840	
	11988	**11986**
11999		
	12116	12110
		12167
	12180	
	12290	12291
12321	12323	12325
	12454	
	12644	
		13448

13651		
	13791ox	13774
13831		**13830**
	13841	
	14004	**14007**
14159	14163	14163
		14309
14863		
	15056	15053
15112		
15143		
15229		
	15262	
		15320
	15333	
15376		15380
	15422	
		15444
15460		
	15470	15478
	15617	
15686	15687	15672
		15948
	16285	
16646		
16762		
	16860ox	16843
	16952	16959
17097		
	17107	17114
	17266	17260
17652	17650	
		17672
	17817	
	17898	17903
		17929
18334	18343	
	18457	
		18470
		18928
	18962	
		19701
	19722	

3.3.2.4 Profiling of Protein Expression in Squamous Carcinoma Cells by RP HPLC offline MALDI MS

The goal in this initial work was to view the changes in protein expression of SCC9 cells transfected with Pg and E-cad since these proteins were lacking in the transformed cells. The comparison of protein expression of SCC9, SCC9-E-cad, and SCC9-Pg cells is portrayed in a mass table in Table 3.6 and in a Venn diagram in Figure 3.10. With external mass calibration, the mass measurement accuracy is normally better than 0.05%. Only those proteins with m/z values within 0.05% error margin were deemed as potentially having the same identity in SCC9, SCC9-E-cad, and SCC9-Pg cells and are represented along the same row in the mass table. Using HPLC offline MALDI MS, 60 proteins for SCC9, 37 proteins for SCC9-E-cad, and 56 proteins for SCC9-Pg cells were detected and are listed by their m/z values in Table 3.5. The number of proteins detected by HPLC offline MALDI MS that were also detected by direct MALDI MS are 7 out of 8 proteins for SCC9, 3 out of 6 proteins for SCC9-E-cad, and 8 out of 12 proteins for SCC9-Pg cells (compare Table 3.2 and Table 3.5). Therefore, the additional proteins detected using HPLC offline MALDI MS compared to the direct MALDI MS method are 53 proteins for SCC9, 34 proteins for SCC9-E-cad, and 48 proteins for SCC9-Pg cells. The total number of proteins detected by both methods are 68 proteins for SCC9, 43 proteins for SCC9-E-cad, and 68 proteins for SCC9-Pg cells.

Table 3.6 Protein profiling of SCC9, SCC9-E-cad, and SCC9-Pg cells of the Human squamous carcinoma of the tongue by RP HPLC offline MALDI MS.

SCC9	SCC9-E-cad	SCC9-Pg
$[M+H]^+$	$[M+H]^+$	$[M+H]^+$
2980	2978	2979
3890	3889	3889
4053		4053
4140	4139	4140
		4199
	4675	4676
		4785
4793		
4939	4937	4938
4966	4964	4965
5357	5352	**5355**
	5515	
5550		
	5629	5630
6070		
	6175	
		6217
6222		
6275	6273	
	6300	
		6306
6400		
		6440
6459		
6512	6511	6513
6581		
6649	**6648**	6649
	6711	
		6762
		6780
		6922
		7176
		7446
7883		
		8095
8181	8179	8182
8239		
8404		
8448	8448	8451

8563	8563	8565
8959		
9151		9153
9261	9261	9260
9313		
		9369
9413		
9514		
9533	9528	9536
9568		
10109	**10099**	**10104**
10200		
	10283	10279
		10346
10518		
	10531	**10526**
10613		10613
10748		
		10826
10836		
		10844
10896		
		10948
11152		
11298		
		11305
		11317
	11324	
		11376
11652		
	11668	
	11729	11725
11751		
11832	11825	11830
	11986	11988
11999		
	12177	12172
	12298	
12321		
		12351
		12463
13651		
		13762
	13783	
13831		

		13849
	14004	14000
	14016	14013
	14027	14023
14159		
		14293
14863		
15112		
15143		
15229		
		15336
	15347	
15376		
15460		
15686		15693
	15834	
16646		
16762		
17097		
		17185
	17215	
17652		
		17883
18334		

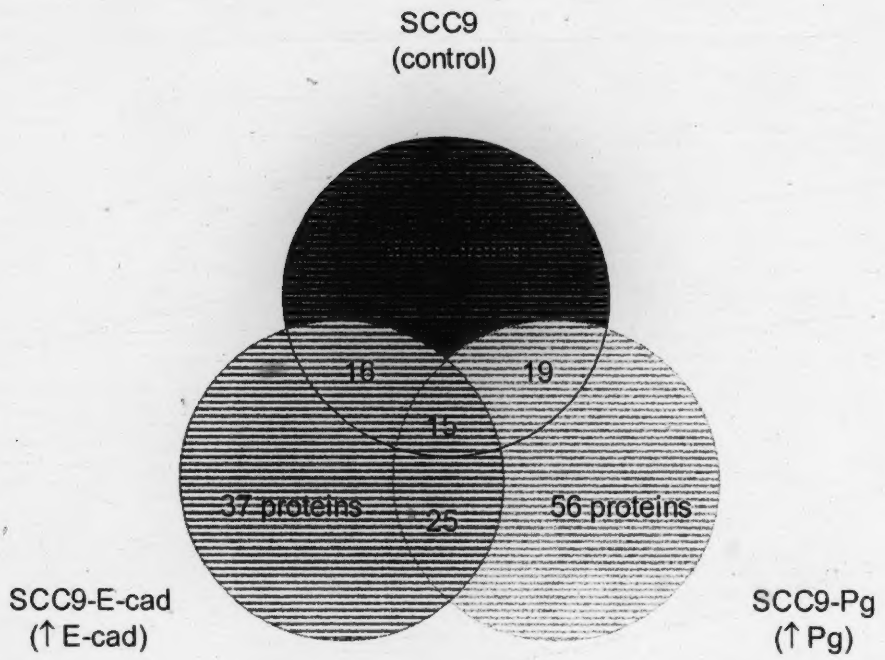


Figure 3.10 Venn diagram of protein profiling of SCC9, SCC9-E-cad and SCC9-Pg cells of the Human squamous carcinoma of the tongue by HPLC offline MALDI MS.

The proteins that are commonly expressed between SCC9, SCC9-E-cad, and SCC9-Pg cells are displayed in the Venn diagram in Figure 3.10. When E-cad expression was introduced in SCC9 cells, a drastic decrease in the amount of proteins was observed in the SCC9-E-cad cells. In total, 37 proteins in SCC9-E-cad cells are detected by HPLC offline MALDI MS compared to 60 proteins in SCC9 cells, and among which 16 proteins are mutually observed. While the introduction of Pg expression produced a similar amount of proteins detected in the resulting SCC9-Pg cells compared to SCC9 cells, only 19 proteins are commonly expressed in both cells. The introduction of E-cad and Pg expression produced significant changes in the protein expression in SCC9-E-cad and SCC9-Pg cells. The variation in the proteins detected as compared to SCC9 cells may suggest the presence of biomarkers that are involved in tumor-suppressive effects leading to the transition from cancer (fibroblastoid) to normal (epidermoid) morphology observed in these cells.[15,16] The initial protein profiling in this work holds great promise for the identification of the differentially expressed proteins and investigation of potential biomarkers in the future work.

The adhesion of cells to their neighbors is important for maintaining tissue architecture and cell polarity and regulating major cellular processes that are essential for motility, growth, differentiation, and survival.[20] The disruption of normal cell-cell adhesion in transformed cells may result from down regulating the expression of cadherin or catenin cell adhesion molecules or by activation of signalling pathways that prevent AJ assembly.[20] The control SCC9 is a Pg and E-cad deficient squamous carcinoma cell line. SCC9 cells express α- and β-catenins and N-cadherin but do not assemble desmosomes and display typical fibroblastic morphology of transformed cells.[15,16] The transformed phenotype of SCC9 cells could result from decreased adhesive properties due to the absence of E-cad or Pg.[16] As observed in SCC9 cells, loss of E-cad expression eliminates AJ formation while re-establishment of AJs in cancer cells by restoration of cadherin expression, as observed in SCC9-E-cad cells, exerts tumor-suppressive effects, including decreased proliferation and motility. The cells became flatter, formed tight epithelial monolayers and showed growth and adhesive properties typical of epidermoid morphology. The transition of 'transformed' to 'normal' phenotype coincides with downregulation of the endogenous N-cadherin and increased synthesis and stability of the catenins. The introduction of Pg expression in SCC9-Pg cells enabled desmosome assembly and increased adhesivity and similarly, induced a fibroblast to epidermoid transition.[16] However, unlike the effects of introduction of E-cad expression, these changes from introduction of Pg corresponded with increased level and stability of N-cadherin and decreased level and stability of β-catenin without any significant effects on α-catenin.

The opposite findings in N-cadherin and β-catenin levels from the introduction of E-cad and Pg expression that both led to the transition to normal morphology of the transformed cells suggests multiple signalling pathways promoting cell adhesion. The introduction of E-cad expression could be acting directly by activating an E-cad-specific signalling pathway, or indirectly by disrupting an N-cad-specific signalling pathway. N-cadherin has similar properties of other classical cadherins, such as solubility, subcellular localization and interactions with α- and β-catenin. The development of the transformed phenotype may result from mutations in N-cadherin of SCC9 cells that could hinder its binding to other cadherins or its interactions with the catenins for normal AJ assembly, whereas, reintroduction of a functional cadherin, E-cad, can restore normal cell adhesion.[15] Evidence for a possible reason behind the increased adhesivity in the induced Pg-expressing cells point towards the reduction of β-catenin levels, which indicates the presence of a functional APC pathway in SCC9 cells. The identification of the differentially expressed proteins may provide insight into what proteins and mechanisms are involved in the cadherin-catenin complexes that function as suppressors of invasion by regulating the integrity and organization of epithelial tissues.

The interaction between the cytoplasmic tail of cadherins with the β-catenins and subsequent interactions of β-catenins with α-catenins and the α-catenins with the actin cytoskeleton is critical for the establishment of stable and functional AJs. Catenin mutations that prevent this interaction have been reported in different types of cancer.[21] Reduced α-catenin expression correlated significantly with a poor prognosis in patients with esophageal squamous cell carcinomas.[30] A homozygous deletion of the α-catenin gene, CTNNA1, leading to loss of cell-cell adhesion, has been found in a human lung cancer cell line.[31] The introduction of wild-type α-catenin into these cells restores normal adhesion. The regulation of α-catenin is critically important in AJ assembly and should be studied in the future work.

3.4 Conclusions

In this initial work, the methods of 2D gel electrophoresis with MALDI MS and RP HPLC offline MALDI MS were used to profile proteins extracted from squamous carcinoma cells and the results indicate that they are powerful methods for profiling expressed proteins in tumor cells. Sequence information of the differentially expressed proteins and biomarkers in SCC9, SCC9-E-cad, and SCC9-Pg cells, may be investigated further in the future work using tandem mass spectrometry. In order to identify the protein, its sequence must be present in the database. Only a small fraction of known human proteins have been sequenced and therefore the identity of expressed oncoproteins, even if they exist in the extract, may be difficult. Therefore, differential protein profiles could serve as biomarkers themselves, whereby different cell lines may be compared on a global level to assess changes in a disease state or to distinguish specific states during disease progression. With sufficient separation of the complex lysate, MALDI MS may prove to be an important analytical tool for detection and profiling of oncoproteins in different cell lysates as well as an important diagnostic tool for monitoring changes in the proteome.

It is possible that the biomarkers have a higher molecular weight and/or different solubility than employed in this study. The extraction method should simplify the challenge of profiling the components of such a complex mixture for proteomics studies. Usually acid or detergent is used to extract tissue samples to yield a complex mixture of proteins of various size and abundance.[28] A more extensive solubilization method and sample preparation should be investigated in the future work.[32] To generate a mass table which is a better represention of the protein components in the cell extracts, the results from an ESI MS experiment may be combined for complementary information.

One of the ~~most~~ critical components in proteomics is sample preparation. This is important because it may affect reproducibility, from the time of sample collection to the point when proteins are introduced for analysis, multiple factors come into play. Efforts should be made to process all the samples in as similar a manner as possible.

Improvements in the sample preparation are necessary to reveal more less abundant proteins. The challenge in the detection of low abundance proteins arise from the required dynamic range of proteins in biological systems, in the ppm range or lower.[3] Preventing sample loss during the various steps, such as during the desalting step, may produce a proteome map better representing the total protein in the extract. Membrane proteins are diverse and important in many functions such as the uptake of nutrients, removal of waste, serve as receptors, and involved in complex degradative and biosynthetic reactions essential for cell existence. The aqueous conditions employed in LC and ESI limit the detection of peptides from membrane proteins, which are currently underrepresented in global proteomic studies. Therefore, a more detail look at the potential membrane proteins which may be lost during the sample handling, may provide the evidence for potential biomarkers in the extract. Efforts to improve separation and unambiguous identification of proteins in the future work can be labour intensive. However, the procedure could be automated to help simplify the investigation, for example, by performing an automated 2D HPLC combined with MS/MS experiment.

3.5 Literature Cited

(1) Fields, S.; *Science* **2001**, *291*, 1221-1224.

(2) Van Eyk, J. E. *Curr. Opin. Mol. Ther* **2001**, *3*, 546-553.

(3) Srinivas, P. R.; Verma, M.; Zhao, Y.; and Srivastava, S. *Clin. Chemistry* **2002**, *48*, 1160-1169.

(4) Gorg, A. et al. *Electrophoresis* **2000**, *21*, 1037-1053.

(5) Hanash, S. M. *Electrophoresis* **2000**, *21*, 1202-1209.

(6) Wulfkuhle, J. D.; Liotta, L. A.; Petricoin, E. F. *Nature Rev. Cancer* **2003**, *3*, 267-275 and references 11-19 therein.

(7) Wu, W.; Hu, W.; Kavanagh, J. J. *Int. J. Gynecol. Cancer* **2002**, *12*, 409-423.

(8) Yates, J. III. *J. Mass Spectrom.* **1998**, *33*, 1-19.

(9) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437-473.

(10) Gygi, S.; Han, D. K.; Gingras, A. C.; Sonenberg, N.; Aebersold, R. *Electrophoresis* **1999**, *20*, 310-319.

(11) Wolters, D. A.; Washburn, M. P.; Yates, J. R. III. *Anal. Chem.* **2001**, *73*, 5683-5690.

(12) Washburn, M. P.; Wolters, D.; Yates, J. R. III. *Nat. Biotechnol.* **2001**, *9*, 242-247.

(13) Hutchens, T. W.; Yip, T. T. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 576-580.

(14) Merchant, M.; Weinberger, S. R. *Electrophoresis* **2000**, *21*, 164-177.

(15) Li, Z.; Gallin, W. J.; Lauzon, G.; Pasdar, M. *J. Cell Sci.* **1998**, *111*, 1005-1019.

(16) Parker, H. R.; Li, Z.; Sheinin, H.; Lauzon, G.; Pasdar, M. *Cell Motil. Cytoskeleton* **1998**, *40*, 87-100.

(17) Coman, D. R. *Cancer Res.* **1944**, *4*, 625-629.

(18) McCutcheon, M.; Coman, D. R.; Moore, F. B. *Cancer* **1948**, *1*, 460-467.

(19) Liotta, L. A.; Steeg, P. S.; Stetler-Stevenson, W. G. *Cell* **1991**, *64*, 327-336.

(20) Conacci-Sorrell, M.; Zhurinsky, J.; and Ben-Ze'ev, A. *J. Clin. Invest.* **2002**, *109*, 987-991.

(21) Hirohashi, S. *Am. J. Pathol.* **1998**, 153, 333-339.

(22) Vleminckx, K.; Vakaet, L.; Mareel, M.; Fiers, W.; van Roy, F. *Cell* **1991**, *66*, 107-119.

(23) Hakimelahi, S.; Parker, H. R.; Gilchrist, A. J.; Barry, M.; Li, Z.; Bleackley, R. C.; Pasdar, M. *J. of Biol. Chem.* **2000**, *275*, 10905-10911.

(24) Nagafuchi, A.; Ishihara, S.; Tsukita, S. *J. Cell Biol.* **1994**, *127*, 235-245.

(25) Gumbiner, B. *Cell* **1996**, *84*, 345-357.

(26) Bradford, M. M. *Anal. Biochem.* **1976**, *72*, 248-254.

(27) Wang, Z.; Dunlop, K.; Long, S. R.; Li, L. *Anal. Chem.* **2002**, *74*, 3174-3182.

(28) Kline, T. R.; Pang, J.; Hefta, S. A.; Opiteck, G. J.; Kiefer, S. E.; Scheffler, J. E. *Anal. Biochem.* **2003**, 315, 183-188.

(29) Piszkiewicz, D.; Landon, M.; Smith, E. L. *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1173-1178.

(30) Nakanishi, Y.; Ochiai, A.; Akimoto, S.; Kato, H.; Watanabe, H.; Tachimori, Y.; Yamamoto, S.; Hirohashi, S. *Oncology* **1997**, *54*, 158-165.

(31) Hirano, S.; Kimoto, N.; Shimoyama, Y.; Hirohashi, S.; Takeichi, M. *Cell* **1992**, *70*, 293-301.

(32) Chong, B. E.; Lubman, D. M.; Rosenspire, A.; Miller, F. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1986-1993.

Chapter 4

Conclusions and Future Work

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry combined with separation techniques remains to be a powerful method for proteomic analysis. For reliable protein identification, peptide mass mapping using MALDI MS in conjunction with MS/MS for sequence information are necessary. The sufficient separation of the complex lysate, including methods of digestion, polyacrylamide gel electrophoresis and liquid chromatography, is equally important for detection and profiling of protein expression.

In Chapter 2, in-gel CNBr cleavage followed by trypsin digestion was demonstrated to be useful for peptide mapping of hydrophobic proteins. Multiple contributing factors are involved in peptide recovery from an in-gel experiment, including digestion efficiency, peptide extraction efficiency, sample loss during sample workup, and ion suppression in MS detection, which are especially critical for analysis of hydrophobic and low abundance proteins. The rationale behind the order of digestion in this sequential digestion method was that the chemical cleaving reagent such as CNBr is expected to be more readily interacting with the hydrophobic sites of a membrane protein in gel, compared to a protease which is bulky and must retain its optimal conformation for activity. After CNBr digestion, the smaller fragments are expected to be more accessible to trypsin for further digestion. The data presented appears to support this hypothesis. The smaller peptides generated also have enhanced efficiency to be extracted from the gel by the passive-elution method.

This method was optimized to detect low amounts of membrane proteins that can be visualized in gel by Coomassie staining method. The use of *n*-octyl-*β*-D-glucopyranoside in the digestion buffer can aid in the solubilization of hydrophobic peptides and improve the peptide signals. Bacteriorhodopsin (BR), a model hydrophobic membrane protein, can be identified from as low as 8 pmol starting sample, compared to nanomolar amounts often used in method developments toward membrane protein analysis by other groups reported in the literature. It was demonstrated that the sequence coverage obtained for BR, as well as the recovery of peptides within the transmembrane region was significantly improved using the CNBr/trypsin digestion method. The sub-microgram detection is adequate for many proteomics application and is comparable to the detection limit of using Coomassie staining of membrane proteins, in which a distinct band is visible with the naked eye.

The application of this method towards nitrate reductase 1 gamma chain (NAR I *E. coli*), a recombinant penta-transmembrane protein and a complex protein mixture extracted from the endoplasmic recticulum (ER) membrane of mouse liver, verified its vital role in detection and subsequent identification of hydrophobic proteins that were not easily characterized from trypsin-only digested peptides. The examples illustrate that, for membrane proteins that can be digested by trypsin-only, the CNBr/trypsin digestion method can provide additional peptides for the purpose of protein identification. The processes involved in data interpretation were presented for the results for Band U7, a gel separated ~70 kDa band from the urea-insoluble fraction of the ER membrane. The use of CNBr/trypsin digestion has allowed the identification and confirmation of more proteins than otherwise possible. It was possible to identify the ATP-binding protein with greater confidence with the aid of CNBr/trypsin cleavage.

The data presented also illustrates limitations of the MALDI MS and MS/MS technique and the importance in obtaining complementary sequence information when a complementary technique such as ESI MS was also used for protein identification of a complex lysate. A number of peptide ions detected in both trypsin and CNBr/trypsin digests of ER membrane proteins were selected for MALDI MS/MS. However, searching these MS/MS spectra against the database did not identify the proteins with high confidence (i.e., each spectrum produced a number of possible candidates with similar low scores in MASCOT). Generally, the MALDI-QqTOF instrument can reliably identify all of the components in a mixture that contains only a few components whose relative abundances differ by a factor of ~10. More complex mixtures may be investigated by LC/ESI-MS with on-line MS/MS capabilities. The complementary set of peptides detected and identified by ESI MS/MS enabled improved sequence coverage for greater confidence of the identified protein and allowed identification of proteins that was not otherwise possible.

In Chapter 3, expression proteomics was applied to the cells of human tongue squamous carcinoma cell line, which has been shown to have decreased adhesive properties compared to normal cells. The purpose of this initial work is to look at the changes in protein expression of SCC9 cells transfected with Pg and E-cad since these proteins were lacking in the transformed cells. The proteomic profiling of these transformed cells has not been previously investigated in the literature. The study is used to obtain a better understanding of the cadherin-catenin system in the regulation of cell proliferation, invasion, and intracellular signalling during cancer.

The protein profiling by three analysis methods were compared: 2D gel electrophoresis, direct MALDI MS and HPLC fractionation with offline MALDI MS. The traditional method of 2D-gel electrophoresis is still a valuable tool for diagnostic purposes and monitoring of distinct protein changes in different cell states. Direct MALDI MS was used to preview the total cell extract in the cancer cells. HPLC prior to MS analysis was used to simplify the complex extract by reducing ion suppression effects and the detected low mass proteins were catalogued in a mass table. LC fractionation used to simplify the complex extract mixture prior to MS analysis is clearly an advantageous step in protein profiling. The resulting mass spectra are much improved in the signal-to-noise ratio, resolution and mass accuracy and also more peaks are detected in the extract compared to direct MALDI MS. The fractionation reduces the number of proteins present in the resulting MALDI sample spot, minimizes effects of ion suppression from more ionisable proteins and allows for a higher probability for each protein to be detected. The data presented from the method of HPLC offline MALDI MS is more reproducible in the lower mass range, < 10,000 Da. Although, less reproducibility is observed in the high mass range, the data can still be credited as reproducible since problems in the detection of high mass proteins is inevitable in MS. The peaks in the mass spectra in this range are generally broader and give rise to poorer mass accuracy. Therefore, future works should include improvements towards the detection of higher masses.

The initial protein profiling in this work holds great promise for the identification of the differentially expressed proteins and investigation of potential biomarkers. Sequence information of the differentially expressed proteins and biomarkers in SCC9, SCC9-E-cad, and SCC9-Pg cells, may be investigated further using tandem mass spectrometry. To be identified, the protein sequence must be present in the database. Only a small fraction of known human proteins have been sequenced and therefore the identity of expressed oncoproteins, even if they exist in the extract, may be difficult. Therefore, differential protein profiles could serve as biomarkers themselves, whereby different cell lines may be compared on a global level to assess changes in a disease state or to distinguish specific states during disease progression. The direct comparisons of protein expression and detection of the differentially expressed proteins by the methods employed in this work were demonstrated to be well suited for biomarker discovery.

It is possible that the biomarkers have a higher molecular weight and/or different solubility than employed in this study. The choice of extraction method should simplify the challenge of profiling the components of such a complex mixture but should also facilitate the results to be a true representative of the proteome of the system. A more detailed look at the potential membrane proteins and low abundance proteins, which may be lost during the sample preparation method employed, may provide the evidence of potential biomarkers in the extract. To generate a mass table, which is a better represention of the protein components in the cell extracts, the results from an ESI MS experiment may be combined for complementary information. Efforts to improve separation and unambiguous identification of proteins in the future work can be labour intensive. However, the procedure could be automated to help simplify the investigation, for example, by performing an automated 2D HPLC combined with MS/MS experiment. Finally the LC MALDI work described in Chapter 3 was limited to qualitative comparison of protein expression profiles from different cell lines, thus, the development of methods for quantitative analysis of proteomes will be required.